PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/12, 15/54, 9/12
// C12Q 1/68, C12N 15/11

A1

(11) International Publication Number: WO 93/15201

(43) International Publication Date: 5 August 1993 (05.08.93)

(21) International Application Number: PCT/US93/00586

(22) International Filing Date: 22 January 1993 (22.01.93)

(30) Priority data:
826,935 22 January 1992 (22.01.92) US

(60) Parent Application or Grant
(63) Related by Continuation
US 826,935 (CIP)
Filed on 22 January 1992 (22.01.92)

(71) Applicant *(for all designated States except US)*: NEW ENGLAND DEACONESS HOSPITAL [US/US]; 185 Pilgrim Road, Boston, MA 02215 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : AVRAHAM, Hava [IL/US]; 50 Radmor Road, Brighton, MA 02135 (US). GROOPMAN, Jerome [US/US]; 79 Druce Street, Brookline, MA 02146 (US). COWLEY, Sally [GB/GB]; 28 Avonmore Road, London W14 8RS (GB). SCADDEN, David [US/US]; 62 Lexington Street, Weston, MA 02193 (US).

(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
With amended claims.

Date of publication of the amended claims: 30 September 1993 (30.09.93)

(54) Title: NOVEL PROTEIN TYROSINE KINASES

(57) Abstract

The identification and isolation of novel protein tyrosine kinase genes present on human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SK	Slovak Republic
CI	Côte d'Ivoire	KZ	Kazakhstan	SN	Senegal
CM	Cameroon	LI	Liechtenstein	SU	Soviet Union
CS	Czechoslovakia	LK	Sri Lanka	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	UA	Ukraine
DK	Denmark	MG	Madagascar	US	United States of America
ES	Spain	ML	Mali	VN	Viet Nam
FI	Finland	MN	Mongolia		

NOVEL PROTEIN TYROSINE KINASES

Description

Background of the Invention

Transduction of signals that regulate cell growth and differentiation is regulated in part by phosphorylation of various cellular proteins. Protein tyrosine kinases are enzymes that catalyze this process. Moreover, many act as growth factor receptors.

Summary of the Invention

The present invention relates to novel protein tyrosine kinase genes present in human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.

The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. The pTK genes have been shown to be present in both megakaryocytic and lymphocytic cells.

The pTK genes of the present invention show significant sequence homology with members of the c-kit subgroup of growth factor receptors with protein tyrosine kinase activity. The c-kit subgroup of receptor tyrosine kinases catalyze the phosphorylation of exogenous substrates, as well as tyrosine residues within their own polypeptide chains. (Ullrich, A. and Schlessinger, J., Cell, 61:203 (1990)). Members of the c-kit subgroup include FLT/FLK (Fetal Liver Kinase), FGF (Fibroblast

Growth Factor Receptor) and NGF (Nerve Growth Factor Receptor).

In particular, fourteen pTK genes have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor tyrosine kinase) were identified in megakaryocytic cells. Five pTK genes, referred to as LpTKs, were identified in lymphocytic cells and have been shown to be present in megakaryocytes as well. One pTK gene, referred to a HpTKs, was identified in human hepatoma cells. Six pTK genes, referred to as bpTK genes, found in human brain tissue.

SAL-S1 is related to the FLT/FLK family of pTKs. SAL-D4 is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3) is related to the NGF receptor family of pTKs.

The pTK genes, which are the subject of the present invention, were identified using two sets of degenerative oligonucleotide primers: a first set which amplifies all pTK DNA segments (SEQ ID NOS:1-2), and a second set which amplifies highly conserved sequences present in the catalytic domain of the c-kit subgroup of pTKs (SEQ ID NOS:3-4). The pTK genes identified in this manner are described below.

SAL-S1 is expressed in several megakaryocytic cell lines, but not in erythroid cell lines. The nucleotide sequence of SAL-S1 was obtained, revealing a sequence containing 158 base pairs. (SEQ ID NO:5). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:6) which exhibited significant sequence homology with known protein tyrosine kinases of the FLT/FLK family. The full

length gene sequence (SEQ ID NO: 17) contains 6827 b.p. and the deduced amino acid sequence (SEQ ID NO: 18) contains 349 residues.

SAL-D4, also expressed in megakaryocytic cells, is a DNA fragment containing the nucleotide sequence of 141 base pairs. (SEQ ID NO:7). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:8) which exhibited significant sequence homology with known protein tyrosine kinases of the FGF receptor family.

The LpTKs, including LpTK 2, LpTK 3, LpTK 4, and LpTK 13 and LpTK 25, are expressed in lymphocytic cells, as well as megakaryocytic cells. The nucleotide sequence (151 base pairs) of the LpTK 3 gene was obtained (SEQ ID NO:11), and exhibited significant homology with known protein tyrosine kinases of the NGF receptor family. The nucleotide sequences of the LpTK 2, LpTK 4, and LpTK 13 genes contained 149 base pairs (SEQ ID NO:9), 137 base pairs (SEQ ID NO:13), and 211 base pairs (SEQ ID NO:15) respectively. LpTK 25 has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). A full length gene sequence has been obtained for LpTK 2 (SEQ ID NO: 19) which contains 7606 b.p. Additional sequencing of LpTK 4 revealed a sequence of 404 b.p. (SEQ ID NO: 21).

The HpTK 5 gene, expressed in human hepatoma cells, has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). Nucleotide sequences of the bpTK's, including bpTK 1, bpTK 2, bpTK 3, bpTK 4, bpTK 5 and bpTK 7 are expressed in human brain tissue encode proteins having the amino acid sequences of SEQ ID NOS: 25-30 respectively.

Thus the present invention includes DNA isolated from a human megakaryocytic cell line, which hybridizes to a

DNA fragment which hybridizes to DNA encoding an amino acid sequence which is highly conserved in the catalytic domain of protein tyrosine kinases of the c-kit subgroup.

The present invention also includes the proteins encoded by the pTK genes identified as described herein, which exhibit significant sequence homology with members of the c-kit subgroup of pTks (i.e. FLT/FLK (SAL-S1), FGF receptor (SAL-D4) or NGF receptor (LpTKS)) as well as the proteins encoded by HpTK 5 and the bpTKs. The present invention also includes SAL-S1, SAL-D4, and LpTK, HpTK and bpTK homologues or equivalents (i.e., proteins which have amino acid sequences substantially similar, but not identical, to that of SAL-S1, SAL-D4, the LpTKs HpTK and the bpTKs, which exhibit tyrosine kinase activity.) This invention further includes peptides (SAL-S1, SAL-D4, LpTK, HpTK and bpTK fragments) which retain tyrosine kinase activity, yet are less than the entire SAL-S1, SAL-D4, LpTK, HpTK and bpTK sequences), monoclonal and polyclonal antibodies specific for SAL-S1, SAL-D4, the LpTKs, HpTK and the bpTKs, and uses for the SAL-S1, SAL-D4, the LpTK, HpTK and the bpTK nucleic acid sequences and SAL-S1, SAL-D4, LpTK, HpTK and bpTK equivalents.

The present invention further includes nucleic acid sequences which hybridize with DNA or RNA encoding the proteins described herein, which exhibit significant sequence homology with the FLT/FLK, FGF receptor or NGF receptor family of protein tyrosine kinases contained within the c-kit subgroup. Such nucleic acid sequences are useful as probes to identify pTK genes in other vertebrates, particularly mammals, and in other cell types. They can also be used as anti-sense

oligonucleotides to inhibit protein tyrosine kinase activity, both in vitro and in vivo.

The SAL-S1, SAL-D4, LpTK, HpTK and bpTK, tyrosine kinases of the present invention can be used as target proteins in conjunction with the development of drugs and therapeutics to modulate cell growth, differentiation and other metabolic functions. The SAL-S1, SAL-D4, LpTK, HpTK or bpTK proteins can be used as agonists or antagonists to other tyrosine kinases. The SAL-S1, SAL-D4, LpTK, HpTK or bpTK tyrosine kinases can also be instrumental in the modulation of megakaryocyte and/or platelet adhesion interactions.

In addition, the SAL-S1, SAL-D4, LpTK, HpTK and bpTK tyrosine kinases can be used in screening assays to detect cellular growth and/or differentiation factors. Using standard laboratory techniques, the ligands of the pTKs of the present invention can be identified. Once identified, assays can be designed to detect these ligands present endogenously, within cells, as well as exogenously, in extra cellular fluids. Assays can also be designed as diagnostic aids to detect these ligands in body fluids such as blood and urine.

Brief Description of the Drawings

Figure 1 depicts the nucleotide sequence of SAL-S1 (SEQ ID NO: 5) and the deduced amino acid sequence (SEQ ID NO:6).

Figure 2 depicts the nucleotide sequence of SAL-D4 (SEQ ID NO:7) and its deduced amino acid sequence (SEQ ID NO:8).

Figure 3A depicts the nucleotide sequence (SEQ ID NO:9) and its deduced amino acid sequence (SEQ ID NO:10) for LpTK 2.

Figure 3B depicts the nucleotide sequence (SEQ ID NO:11) and its deduced amino acid sequence (SEQ ID NO:12) for LpTK 3.

Figure 3C depicts the nucleotide sequence (SEQ ID NO:13) and its deduced amino acid sequence (SEQ ID NO:14) for LpTK 4.

Figure 3D depicts the nucleotide sequence (SEQ ID NO:15) and its deduced amino acid sequence (SEQ ID NO:16) for the LpTK 13.

Figure 4A-4J depicts the full-length nucleotide sequence (SEQ ID NO: 17) and its deduced amino acid sequence (SEQ ID NO: 18) for SAL-S1.

Figure 5A-5J depicts the full length nucleotide sequence (SEQ ID NO: 19) and the deduced amino acid sequence (SEQ ID NO: 20) for LpTK2.

Figure 6 depicts the partial nucleotide sequence (SEQ ID NO: 21) for LpTK4.

Figure 7A-7D depicts the full length nucleotide sequence (SEQ ID NO: 22) for LpTK25.

Figure 8A-8F depicts the full length nucleotide sequence (SEQ ID NO: 23) and the deduced amino acid sequence (SEQ ID NO: 24) for HpTK5.

Figure 9 depicts the amino acid sequence (SEQ ID NO: 25) of bpTK1.

Figure 10 depicts the amino acid sequence (SEQ ID NO: 26) of bpTK2.

Figure 11 depicts the amino acid sequence (SEQ ID NO: 27) of bpTK3.

Figure 12 depicts the amino acid sequence (SEQ ID NO: 28) of bpTK4.

Figure 13 depicts the amino acid sequence (SEQ ID NO: 29) of bpTK5.

Figure 14 depicts the amino acid sequence (SEQ ID NO: 30) of bpTK7.

Detailed Description of the Invention

Novel protein tyrosine kinase genes have been identified, their nucleic acid sequences determined, and the amino acid sequences of the encoded proteins deduced. The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology to with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. These genes have been shown to be present in both megakaryocytic and lymphocytic cells.

To facilitate the isolation and identification of these novel pTKs, two sets of DNA probes were used, as described in the Exemplification. The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2) (Matthews, W. Cell 65: 1143 (1991; Wilks, A. F. Proc. Natl. Acad. Sci. USA 86:1603 (1989)). These sequences were used as primers in a polymerase chain reaction to amplify tyrosine kinase DNA segments. (Mullis, K. et al., Cold Spring Harbor Symp. Advan. Biol. 51:263 (1986).

The second set consisted of two oligonucleotide sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4)

designed to amplify the nucleic acid sequence which encodes the highly conserved regions of the catalytic domains of the c-kit family of protein tyrosine kinases. These sequences were used as primers in the polymerase chain reaction in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK primers were identified, isolated and subsequently sequenced.

In particular, fourteen pTK genes exhibiting significant homology with the c-kit subgroup of protein tyrosine kinases have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor) were identified in several megakaryocytic cell lines, including CMK 11-5, DAMI, UT-7 and UT-7 grown in erythropoietin, but not in the erythroid cell lines HEL, PMA stimulated HEL cells, or K562. Five pTK genes, referred to as LpTKs, were identified in lymphocytic, as well as in megakaryocytic cells. One pTK gene, referred to as HpTK5 was identified in human hepatoma cells and six genes, referred to as bpTKs, were identified in human brain tissue.

SAL-S1 (SEQ ID NO:6 and 18) encoded by the nucleic acid sequence SEQ ID NOS:5 and 17, exhibits significant homology with the FLT/FLK family of pTKs. SAL-D4 (SEQ ID NO:8) encoded by SEQ ID NO:7, is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3 (SEQ ID NO:12) encoded by the SEQ ID NO:11, is related to the NGF receptor family of pTKs. The remaining LpTKs, LpTK2 (SEQ ID NO:10) encoded by SEQ ID NO:9; LpTK4 (SEQ ID NO:14) encoded by SEQ ID NO:13; LpTK13 (SEQ ID NO:16) encoded by SEQ ID NO:15 LpTK25 encoded by SEQ ID NO: 22, also exhibit

sequence homology with known protein tyrosine kinases (Data not shown).

HpTK5 (SEQ ID NO: 24) encoded by SEQ ID NO: 23 and the bpTKs 1, 2, 3, 4, 5 and 7 (SEQ ID NOS: 25-30 respectively, also exhibit sequence homology with known protein tyrosine kinases.

Thus, as described above, DNA which hybridize with DNA encoding amino acid sequences present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases have been isolated and sequenced. These isolated DNA sequences, collectively referred to as pTKs genes, (and their deduced amino acid sequences) have been shown to exhibit significant sequence homology with known members of receptor tyrosine kinase families.

Once isolated, these DNA fragments can be amplified using known standard techniques such as PCR. These amplified fragments can then be cloned into appropriate cloning vectors and their DNA sequences determined.

These DNA sequences can be excised from the cloning vectors, labeled with a radiolabeled nucleotide such as ^{32}P and used to screen appropriate cDNA libraries to obtain the full-length cDNA clone.

The pTk genes as described above have been isolated from the source in which they occur naturally, i.e. megakaryocyte and lymphocytic cells. The present invention is intended to include pTk genes produced using genetic engineering techniques, such as recombinant technology, as well as pTk genes that are synthesized chemically.

The deduced amino acid sequences of the pTK genes include amino acid sequences which encode peptides exhibiting significant homology with the catalytic domain

of protein tyrosine kinases of the c-kit subgroup of tyrosine kinases. These proteins, encoded by the pTk genes, can include sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence, resulting in a silent change, that is a change not detected phenotypically. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent substitution.

In addition, the protein structure can be modified by deletions, additions, inversion, insertions or substitutions of one or more amino acid residues in the sequence which do not substantially detract from the desired functional tyrosine kinases properties of the peptide.

Modified pTKs of the present invention, with receptor tyrosine kinase activity can be made using recombinant DNA techniques, such as excising it from a vector containing a cDNA encoding such a protein, or by synthesizing DNA encoding the desired protein mechanically and/or chemically using known techniques.

An alternate approach to producing the pTKs of the present invention is to use peptide synthesis to make a peptide or polypeptide having the amino acid sequence of such a protein. The peptides or modified equivalents thereof, can be synthesized directly by standard solid or liquid phase chemistries for peptide synthesis.

Preferably, the pTKs of the present invention will be produced by inserting DNA encoding the proteins into an appropriate vector/host system where it will be expressed.

The DNA sequences can be obtained from sources in which they occur naturally, can be chemically synthesized or can be produced using standard recombinant technology.

This invention also pertains to an expression vector comprising a pTK gene of the present invention, encoding for a protein which exhibits receptor tyrosine kinase activity.

The pTK genes of the present invention can be used for a number of diagnostic and therapeutic purposes. For example, the nucleic acid sequences of the pTK genes can be used as probes to identify other protein tyrosine kinases present in other cell types, including eukaryotic and prokaryotic cell types.

The nucleic acid sequences can be used to design drugs that directly inhibit the kinase activity of protein tyrosine kinases, or to design peptides that bind to the catalytic domain of tyrosine kinases, thus inhibiting their activity. These sequences can also be used to design anti-sense nucleotides that can also inhibit, or destroy, tyrosine kinase activity. Such inhibition of tyrosine kinase activity would be desirable in pathological states where decreased cellular proliferation would be beneficial, such as leukemias or other malignancies.

The nucleic acid sequences can also be used to design drugs, peptides or anti-sense nucleotides as above, but with enhancing, rather than inhibitory effects, on tyrosine kinases. Such enhanced tyrosine kinase activity would result in increasing the phosphorylation of substrates (exogenous, as well as endogenous tyrosine residues). Enhanced effects would be desirable in states

where increased cellular proliferation would be beneficial, such as anemias, bleeding disorders and during surgical procedures.

The pTK genes of the present invention can also be used to obtain soluble fragments of receptor tyrosine kinases, capable of binding their respective ligands (i.e. fibroblast growth factor).

pTK genes encoding soluble receptor tyrosine kinase fragments can be produced using recombinant DNA techniques or synthetically. In either case, the DNA obtained encodes a soluble pTK fragment which lacks a substantial portion of the hydrophobic transmembrane region to permit solubilization of the fragment.

These soluble pTK protein fragments can be introduced exogenously to act as competitors with the endogenous, membrane bound pTK for their respective ligands, thus inhibiting tyrosine kinase activity. Alternately, a modified soluble pTK protein fragment can be introduced which binds the ligand but does not activate kinase activity.

These soluble pTK protein fragments can also be used in binding assays to detect ligands such as growth and differentiation factors. Once these ligands are identified, they may be altered or modified to inhibit or enhance kinase activity. For example, the ligands may be modified or attached to substances that are toxic to the cell, such a ricin, thus destroying the target cell. The substance may be a super-activating substance which, after binding to the pTK, may substantially increase the kinase activity, or activate other growth factors.

pTk genes of the present invention would also be useful to develop diagnostic tools for in vitro screening assays for ligands such as growth factors or differentiation factors that inhibit or enhance kinase activity. The proteins encoded by the pTK genes can also be used in such assays, or as immunogens to produce monoclonal or polyclonal antibodies to be used in such assays.

Such antibodies can also be used in methods of treating conditions in which an individual would benefit therapeutically if protein tyrosine kinase activity could be modified, such as increasing platelet production in bleeding disorders.

The present invention will now be illustrated by the following Exemplification, which is not intended to be limiting in any way.

<u>Exemplification: The Identification and Isolation of the pTK Genes</u>

To facilitate the isolation and identification of these novel pTK genes, two sets of DNA probes were used. (See Table).

The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2(SEQ ID NO:2). These sequences were used as polymerase chain reaction (PCR) primers, using standard PCR techniques, to amplify tyrosine kinase DNA segments.

The second set consisted of two oligonucleotide sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4) selected from the highly conserved regions of the catalytic domains of the c-kit subgroup of protein

tyrosine kinases. These sequences were also used as polymerase chain reaction primers in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK primers were identified, isolated and subsequently sequenced using known laboratory techniques.

TABLE

First Round of Amplification

PTK1
CGGATCCACAGNGACCT

PTK2
GGAATTCCAAAGGACCAGACGTC

Second Round of Amplification

PTK3 (kit family specific)
CGGATCCATCCACAGAGATGT

PTKKW (kit family specific)
GGAATTCCTTCAGGAGCCATCCACTT

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

CLAIMS

The invention claimed is:

1. Isolated DNA of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.

2. Isolated DNA of Claim 1 having a nucleotide sequence selected from the group of nucleotide sequences consisting of:
   a) SAL-S1 (SEQ ID NOS:5 and 7);
   b) SAL-D4 (SEQ ID NO:7);
   c) LpTK 2 (SEQ ID NOS:9 and 19);
   d) LpTK 3 (SEQ ID NO:11);
   e) LpTk 4 (SEQ ID NOS:13 and 21);
   f) LpTK 13 (SEQ ID NO:15);
   g) LpTK 25 (SEQ ID NO: 22 and
   h(HpTK 5 (SEQ ID NO: 23).

3. Isolated DNA of Claim 1 which encodes an amino acid sequence selected from the group consisting of:
   a) SAL-S1 (SEQ ID NOS:6 and 18);
   b) SAL-D4 (SEQ ID NO:8);
   c) LpTK 2 (SEQ ID NOS:10 and 20);
   d) LpTK 3 (SEQ ID NO:12);
   e) LpTK 4 (SEQ ID NO:14);
   f) LpTK 13 (SEQ ID NO:16);
   g) HpTK 5 (SEQ ID NO:24);

h) bpTK 1 (SEQ ID NO:25);
i) bpTK 2 (SEQ ID NO:26);
j) bpTK 3 (SEQ ID NO:27);
k) bpTK 4 (SEQ ID NO:28);
l) bpTK 5 (SEQ ID NO:29); and
m) bpTK 7 (SEQ ID NO:30).

4. Isolated DNA of human megakaryocytic origin which comprises a DNA fragment whose sequence encodes an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.

5. Isolated DNA of Claim 4 which encodes an amino acid sequence selected from the group consisting of:
a) SAL-S1 (SEQ ID NO:6);
b) SAL-D4 (SEQ ID NO:8);
c) LpTK 2 (SEQ ID NO:10);
d) LpTK 3 (SEQ ID NO:12);
e) LpTk 4 (SEQ ID NO:14); and
f) LpTK 13 (SEQ ID NO:16).
g) HpTK 5 (SEQ ID NO:24);
h) bpTK 1 (SEQ ID NO:25);
i) bpTK 2 (SEQ ID NO:26);
j) bpTK 3 (SEQ ID NO:27);
k) bpTK 4 (SEQ ID NO:28);
l) bpTK 5 (SEQ ID NO:29); and
m) bpTK 7 (SEQ ID NO:30).

6. A homogeneous protein of human megakaryocytic origin which includes an amino acid sequence exhibiting sequence homology with the catalytic domain of tyrosine kinases of the c-kit family.

7. A homogeneous protein of Claim 6 in which the amino acid sequence is selected from the group consisting of:

a) SAL-S1 (SEQ ID NO:6);
b) SAL-D4 (SEQ ID NO:8);
c) LpTK 2 (SEQ ID NO:10);
d) LpTK 3 (SEQ ID NO:12);
e) LpTk 4 (SEQ ID NO:14); and
f) LpTK 13 (SEQ ID NO:16).
g) HpTK 5 (SEQ ID NO:24);
h) bpTK 1 (SEQ ID NO:25);
i) bpTK 2 (SEQ ID NO:26);
j) bpTK 3 (SEQ ID NO:27);
k) bpTK 4 (SEQ ID NO:28);
l) bpTK 5 (SEQ ID NO:29); and
m) bpTK 7 (SEQ ID NO:30).

8. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FLT/FLK family of protein tyrosine kinases.

9. A protein of Claim 8 encoded by the nucleotide sequence (SEQ ID NO:5).

10. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO:6).

19. A DNA expression vector containing a DNA sequence of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases.

20. The DNA expression vector of Claim 17 containing a DNA sequence selected from the group consisting of:
a) SAL-S1 (SEQ ID NO:5);
b) SAL-D4 (SEQ ID NO:7);
c) LpTK 2 (SEQ ID NO:9);
d) LpTK 3 (SEQ ID NO:11);
e) LpTk 4 (SEQ ID NO:13); and
f) LpTK 13 (SEQ ID NO:15).
g) LpTK 25 (SEQ ID NO: 22 and
h(HpTK 5 (SEQ ID NO: 23).

21. A cell transformed by the expression vector of Claim 17.

AMENDED CLAIMS
[received by the International Bureau on 31 August 1993 (31.08.93); original claims 1, 2, 4-7 and 14-21 amended; remaining claims unchanged (5 pages)]

1. Isolated DNA of human origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.

2. Isolated DNA of Claim 1 having a nucleotide sequence selected from the group of nucleotide sequences consisting of:
a) SAL-S1 (SEQ ID NOS:5 and 17);
b) SAL-D4 (SEQ ID NO:7);
c) LpTK 2 (SEQ ID NOS:9 and 19);
d) LpTK 3 (SEQ ID NO:11);
e) LpTk 4 (SEQ ID NOS:13 and 21);
f) LpTK 13 (SEQ ID NO:15);
g) LpTK 25 (SEQ ID NO: 22); and
h) HpTK 5 (SEQ ID NO: 23).

3. Isolated DNA of Claim 1 which encodes an amino acid sequence selected from the group consisting of:
a) SAL-S1 (SEQ ID NOS:6 and 18);
b) SAL-D4 (SEQ ID NO:8);
c) LpTK 2 (SEQ ID NOS:10 and 20);
d) LpTK 3 (SEQ ID NO:12);
e) LpTK 4 (SEQ ID NO:14);
f) LpTK 13 (SEQ ID NO:16);
g) HpTK 5 (SEQ ID NO:24);

h) bpTK 1 (SEQ ID NO:25);
i) bpTK 2 (SEQ ID NO:26);
j) bpTK 3 (SEQ ID NO:27);
k) bpTK 4 (SEQ ID NO:28);
l) bpTK 5 (SEQ ID NO:29); and
m) bpTK 7 (SEQ ID NO:30).

4. Isolated DNA of human origin which comprises a DNA fragment whose sequence encodes an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.

5. Isolated DNA of Claim 4 which encodes an amino acid sequence selected from the group consisting of:
a) SAL-S1 (SEQ ID NOS:6 and 18);
b) SAL-D4 (SEQ ID NO:8);
c) LpTK 2 (SEQ ID NOS:10 and 20);
d) LpTK 3 (SEQ ID NO:12);
e) LpTk 4 (SEQ ID NO:14);
f) LpTK 13 (SEQ ID NO:16);
g) HpTK 5 (SEQ ID NO:24);
h) bpTK 1 (SEQ ID NO:25);
i) bpTK 2 (SEQ ID NO:26);
j) bpTK 3 (SEQ ID NO:27);
k) bpTK 4 (SEQ ID NO:28);
l) bpTK 5 (SEQ ID NO:29); and
m) bpTK 7 (SEQ ID NO:30).

6. A homogeneous protein of human origin which includes an amino acid sequence exhibiting sequence homology with the catalytic domain of tyrosine kinases of the c-kit family.

7. A homogeneous protein of Claim 6 in which the amino acid sequence is selected from the group consisting of:

a) SAL-S1 (SEQ ID NOS:6 and 18);
b) SAL-D4 (SEQ ID NO:8);
c) LpTK 2 (SEQ ID NOS:10 and 20);
d) LpTK 3 (SEQ ID NO:12);
e) LpTk 4 (SEQ ID NO:14);
f) LpTK 13 (SEQ ID NO:16);
g) HpTK 5 (SEQ ID NO:24);
h) bpTK 1 (SEQ ID NO:25);
i) bpTK 2 (SEQ ID NO:26);
j) bpTK 3 (SEQ ID NO:27);
k) bpTK 4 (SEQ ID NO:28);
l) bpTK 5 (SEQ ID NO:29); and
m) bpTK 7 (SEQ ID NO:30).

8. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FLT/FLK family of protein tyrosine kinases.

9. A protein of Claim 8 encoded by the nucleotide sequence (SEQ ID NO:5).

10. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO:6).

11. A protein of Claim 8 encoded by the nucleotide sequence (SEQ ID NO:17).

12. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO: 18).

13. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FGF receptor family of protein tyrosine kinases.

14. A protein of Claim 13 encoded by the nucleotide sequence (SEQ ID NO:7).

15. A protein of Claim 13 encoded by the amino acid sequence (SEQ ID NO:8).

16. A protein of human megakaryocytic or lymphocytic origin which exhibits significant sequence homology with the NGF receptor family of protein tyrosine kinases.

17. A protein of Claim 16 encoded by the nucleotide sequence (SEQ ID NO:11).

18. A protein of Claim 16 encoded by the amino acid sequence (SEQ ID NO:12).

19. A DNA expression vector containing a DNA sequence of human origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases.

20. The DNA expression vector of Claim 19 containing a DNA sequence selected from the group consisting of:
a) SAL-S1 (SEQ ID NOS:5 and 17);
b) SAL-D4 (SEQ ID NO:7);
c) LpTK 2 (SEQ ID NOS:9 and 19);
d) LpTK 3 (SEQ ID NO:11);
e) LpTk 4 (SEQ ID NOS:13 and 21);
f) LpTK 13 (SEQ ID NO:15);
g) LpTK 25 (SEQ ID NO: 22); and
h) HpTK 5 (SEQ ID NO: 23).

21. A cell transformed by the expression vector of Claim 19.

sals1 (160 bases)
FLKI-LIKE

```
1        PTKI/3 PRIMERS        21                 41
5'ggatcctgtgcatcagtgacttagggctaggaacattctgctgtcggaaagcgacgtggt
   D  P  V  H  Q  a  L  R  A  R  H  I  L  L  S  E  S  D  V  V

61                            81                 101
gaagatctgtgactttggccttgcccgggacatctacaaagaccccagctacgtccgcaa
   K  I  C  D  F  G  L  A  R  D  I  Y  K  D  P  S  Y  V  R  K

121                           141   PTKKW PRIMER
gcatgcccggctgcccctgaagtggatggcgccagaattc 3'
   H  A  R  L  P  L  K  W  H  A  P  E  F
```

FIGURE 1

sald4 (147 bases)-
FGFR-LIKE

```
   1    PTK1/3 PRIMERS    21                  41
5' ggatccattcacagagacctagcagcacgcaacatcctggtctcagaggacctggtaacc
    G  S  I  H  R  D  L  A  A  R  N  I  L  V  S  E  D  L  V  T

   61                   81                  101
   aaggtcagcgactttggcctggccaaagccgagcggaaggggctagactcaagccggctg
    K  V  S  D  F  G  L  A  K  A  E  R  K  G  L  D  S  S  R  L

   121   PTKKW PRIMER   141
   cccgtcaaatggatggctcccgaattc 3'
    P  V  K  W  M  A  P  E  F
```

FIGURE 2

LpTK2

GTTGGAATTCCTTCCGGCGCCATCCATTTCACCGGCAGCTTTATTTCGTGTCTAGATTCA

TAGATGTCTTCATTATCTACCTTAAAAACTCTGGCAAGTCCAAAATCTGCTACTTTGTAG

ATATTATGTTCACCAACGAGGACATTCCT

FIGURE 3A

LpTK3

GTGCACAGGGATCTCGCGGCTCGGAACATCCTCGTCGGGGAAAACACCCTCTCGAAAGTT

GGGGACTTCGGGTTAGCCAGGCTTATCAAGGAGGACGTCTACCTCTCCCATGACCACAAT

ATCCCCTACAAATGGATGGCCCCTGAGGGAA

FIGURE 3B

LpTK4

GTTCACCGAGATCTCAAGTCCAACAACATTTTGCTGCTGCAGCCCATTGAGAGTGACGAC

ATGGAGCACAAGACCCTGAAGATCACCGACTTTGGCCTGGCCCGAGAGTGGCACAAAACC

ACACAAATGAGTGCCGC

FIGURE 3C

LpTK13

GTCAATCGTGACCTCGCCGCCCCGAAATGTGTTGCTAGTTACCCAACATTACGCCAAGATC

AGTGATTTCGGACTTTCCAAAGCACTGCGTGCTGATGAAAACTACTACAAGGCCCAGACC

CATGGAAAGTGGCCTGTCAAGTGGTACGCTCCGGAATGCATCAACTACTACAAGTTCTCC

AGCAAAAGCGATGTCTGGTCCTTTGGAATTC

FIGURE 3D

```
  1 TTCGAGCTCG CCCGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC
    AAGCTCGAGC GGGCTGTAAC TAATAACTGA TCAATAATTA TCATTAGTTA ATGCCCCAGT AATCAAGTAT CGGGTATATA CCTCAAGGCG CAATGTATTG

101 TTACGGTAAA TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA
    AATGCCATTT ACCGGGCGGA CCGACTGGCG GGTTGCTGGG GGCGGGTAAC TGCAGTTATT ACTGCATACA AGGGTATCAT TGCGGTTATC CCTGAAAGGT

201 TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT
    AACTGCAGTT ACCCACCTCA TAAATGCCAT TTGACGGGTG AACCGTCATG TAGTTCACAT AGTATACGGT TCATGCGGGG GATAACTGCA GTTACTGCCA

301 AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC
    TTTACCGGGC GGACCGTAAT ACGGGTCATG TACTGGAATA CCCTGAAAGG ATGAACCGTC ATGTAGATGC ATAATCAGTA GCGATAATGG TACCACTACG

401 GGTTTTGGCA GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA
    CCAAAACCGT CATGTAGTTA CCCGCACCTA TCGCCAAACT GAGTGCCCCT AAAGGTTCAG AGGTGGGGTA ACTGCAGTTA CCCTCAAACA AAACCGTGGT

501 AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC CCATTGACGC AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT
    TTTAGTTGCC CTGAAAGGTT TTACAGCATT GTTGAGGCGG GGTAACTGCG TTTACCCGCC ATCCGCACAT GCCACCCTCC AGATATATTC GTCTCGAGCA

601 TTAGTGAACC GTCAGATCCC CTGGAGACGC CATCCACCCT GTTTTGACCT CCATAGAAGA CACCGGGACC GATCCAGCCT CCGCGGCCGG GAACGGTGCA
    AATCACTTGG CAGTCTAGCG GACCTCTGCG GTAGGTGCGA CAAAACTGGA GGTATCTTCT GTGGCCCTGG CTAGGTCGGA GGCGCCGGCC CTTGCCACGT
```

FIGURE 4A

```
701  TTGGAACGCG GATTCCCCGT GCCAAGAGTG ACGTAAGTAC CGCCTATAGA GTCTATAGGC CCACTTGGCT TCGTTAGAAC GCGGCTACAA TTAATACATA
     AACCTTGCGC CTAAGGGGCA CGGTTCTCAC TGCATTCATG GCGGATATCT CAGATATCCG GGTGAACCGA AGCAATCTTG CGCCGATGTT AATTATGTAT

801  ACCTTATGTA TCATACACAT ACGATTTAGG TGACACTATA GAATAACATC CACTTTGCCT TTCTCTCCAC AGGTGTCCAC TCCCAGGTCC AACTGCACCT
     TGGAATACAT AGTATGTGTA TGCTAAATCC ACTGTGATAT CTTATTGTAG GTGAAACGGA AAGAGAGGTG TCCACAGGTG AGGGTCCAGG TTGACGTGGA

901  CGGTTCTATC GATTGAATTC CCCGGGGATC CTCTAGAGAT CCCTCGACCT CGAGATCCAT TGTGCTGGCG CGGATTCTTT ATCACTGATA AGTTGGTGGA
     GCCAAGATAG CTAACTTAAG GGGCCCCTAG GAGATCTCTA GGGAGCTGGA GCTCTAGGTA ACACGACCGC GCCTAAGAAA TAGTGACTAT TCAACCACCT

1001 CATATTATGT TTATCAGTGA TAAAGTGTCA AGCATGACAA AGTTGCAGCC GAATACAGTG ATCCGTGCCG CCCTAGACCT GTTGAACGAG GTCGGCGTAG
     GTATAATACA AATAGTCACT ATTTCACAGT TCGTACTGTT TCAACGTCGG CTTATGTCAC TAGGCACGGC GGGATCTGGA CAACTTGCTC CAGCCGCATC

1101 ACGGTCTGAC GACACGCAAA CTGGCGGAAC GGTTGGGGGT TCAGCAGCCG GCGCTTTACT GGCACTTCAG GAACAAGCGG GCGCTGCTCG ACGCACTGGC
     TGCCAGACTG CTGTGCGTTT GACCGCCTTG CCAACCCCCA AGTCGTCGGC CGCGAAATGA CCGTGAAGTC CTTGTTCGCC CGCGACGAGC TGCGTGACCG

1201 CGAAGCCATG CTGGCGGAGA ATCATAGCAC TTCGGTGCCG AGAGCCGACG ACGACTGGCG CTCATTTCTG ACTGGGAATG CCCGCAGCTT CAGGCAGGCG
     GCTTCGGTAC GACCGCCTCT TAGTATCGTG AAGCCACGGC TCTCGGCTGC TGCTGACCGC GAGTAAAGAC TGACCCTTAC GGGCGTCGAA GTCCGTCCGC

1301 CTGCTCGCCT ACCGCCAGCA CAATGGATCT CGAGGGATCT TCCATACCTA CCAGTTCTGC GCCTGCAGGT CGCGGCCGCA CTACTCTTTG ATGTATTACT
     GACGAGCGGA TGGCGGTCGT GTTACCTAGA GCTCCCTAGA AGGTATGGAT GGTCAAGACG CGGACGTCCA GCGCCGGCGT GATGAGAAAC TACATAATGA
```

FIGURE 4B

```
1401 CATATTACCA AGGAATAACT GGCGGGCACA GGGTCAGGTG CTGAAGGGAC ATTGTGAGAA GTGACCTAGA AGGCAAGAGG TGAGCCCTCT GTCACGCTGG
     GTATAATGGT TCCTTATTGA CCGCCCGTGT CCCAGTCCAC GACTTCCCTG TAACACTCTT CACTGGATCT TCCGTTCTCC ACTCGGGAGA CAGTGCGACC

1501 CATAAGGGCC GCTTGAGGGC TCTTTGGTCA AGCAGTAACG CCAGTGTCTG GGAAGGCACC TGTTACTCAG CAGACCATGA AAGGGCGTCT CCCTTTCCTT
     GTATTCCCGG CGAACTCCCG AGAAACCAGT TCGTCATTGC GGTCACAGAC CCTTCCGTGG ACAATGAGTC GTCTGGTACT TTCCCGCAGA GGGAAAGGAA

1601 CGAGCAGTCA GGGAACACTC TGCTCCACCA GCTTCTTGTG GGAGGCTGGA TATTATCCAG GCCTGCCCGC AGTCATCCGG AGGCCTAACC CCTCCCTGTG
     CCTCCTCAGT CCCTTGTGAG ACGAGGTGGT CGAAGAACAC CCTCCGACCT ATAATAGGTC CGGACGGGCG TCAGTAGGCC TCCGGATTGG GGAGGGACAC

1701 GTGCTTCAGT GGTCACACTC CTTGTCCACT TTCATGCTCC TCTTGGCCTC CTGGTTCCTC TTGGAAGTTT GTAGTAGATA GCAGAAGAAA TAGCGAAAGT
     CACGAAGTCA CCAGTGTGAG GAACAGGTGA AAGTACGAGG AGAACCGGAG GACCAAGGAG AACCTTCAAA CATCATCTAT CGTCTTCTTT ATCGCTTTCA

1801 CTTAAAGTCT TTGATCTTTC TTATAAGTGC AGAGAAGAAA TGCTGACGTA TGCTGCCTTC TCTCTCTCTG CTTCAGCTAC CTGAAGCCGC TTTCTTGTCT
     GAATTTCAGA AACTAGAAAG AATATTCACG TCTCTTCTTT ACGACTGCAT ACGACGGAAG AGAGAGAGAC GAAGTCGATG GACTTCGGCG AAAGAACAGA
                                                                                          Q  R P G  S E Q R
349

1901 ATACCTGCTC TCTATCTGCT CACACTCCTC CGAGGCCAGC ACCATCCCAC TGTCTGTCTG GTTGTCCACA GAGCCTTTGT AGGTCGTTGG GGTCATGGGG
     TATGGACGAG AGATAGACGA GTGTGAGGAG GCTCCGGTCG TGGTAGGGTG ACAGACAGAC CAACAGGTGT CTCGGAAACA TCCAGCAACC CCAGTACCCC
341   Y  R  S   E  I  Q   E  C  E  E   S  A  L   V  H  G   S  D  T  Q   N  D  V   S  G  K   Y  T  T  P   T  H  P

2001 AATTCCTCAA ATGTCTTCAT CCTGGAGGAA CCACGGGTCT CAGCCCCTCT GGCCAGGCAC CCGGGAAAGG ACACCCAGTT GTAATACCTG GCGGCCAGCC
     TTAAGGAGTT TACAGAAGTA GGACCTCCTT GGTGCCCAGA GTCGGGGAGA CCGGTCCGTG GGCCCTTTCC TGTGGGTCAA CATTATGGAC CGCCGGTCGG
308   P  E  E   F  T  K  H   R  S  S   G  R  T   E  A  G  R   A  L  C   G  P  F   S  V  W  N   Y  Y  R   A  A  L
```

FIGURE 4C

```
2101 TGTGGCGCTG CAGGCTTGGC GGGCTGTCCT CAGCGTCAGC CTGGGCGATG TGTAGGGCCA TGGTGGACAC CTGCGAGAAG CTGCCCTCTT CTGAGCTCTG
     ACACCGCGAC GTCCGAACCG CCCGACAGGA GTCGCAGTCG GACCCGCTAC ACATCCCGGT ACCACCTGTG GACGCTCTTC GACGGGAGAA GACTCGAGAC
275  S  H  R  Q  L  S  P  P  S  D  E  A  D  A  Q  A  I  H  L  A  H  T  S  V  Q  S  F  S  G  E  E  S  S  Q

2201 AGAGCTGCGC GGGGCCATGC AGACCTCCTC TTCCTCTTGC AGGCCCCTGC CCTGGAGCAG GTCCCCCAGG ATCTCCACCA GCTCCGAGAA TGCAGGTCTC
     TCTCGACGCG CCCCGGTACG TCTGGAGGAG AAGGAGAACG TCCGGGGACG GGACCTCGTC CAGGGGGTCC TAGAGGTGGT CGAGGCTCTT ACGTCCAGAG
241  S  S  R  P  A  M  C  V  E  E  E  E  Q  L  G  R  G  Q  L  L  D  G  L  I  E  V  L  E  S  F  A  P  R

2301 GCCTTGGGGT CTCCGGACCA GCAGTTCAGC ATGATGCGGC GTATGGCGGG AGTGGCCAGC TCCGGGGCCC TCATCCTTGT GCCGTCTCTC AGCCGCTGGC
     CGGAACCCCA GAGGCCTGGT CGTCAAGTCG TACTACGCCG CATACCGCCC TCACCGGTCG AGGCCCCGGG AGTAGGAACA CGGCAGAGAG TCGGCGACCG
208  A  K  P  D  G  S  W  C  N  L  M  I  R  R  I  A  P  T  A  L  E  P  A  R  M  R  T  G  D  R  L  R  Q

2401 AGAACTCCTC ATTGATCTGC ACCCCAGGGT ACGGGGAGGC CCCCAGAGAG AAGATCTCCC AGAGAAGCAC CCCAAAGGAC CACACGTCAC TCTGCGTGGT
     TCTTGAGGAG TAACTAGACG TGGGGTCCCA TGCCCCTCCG GGGGTCTCTC TTCTAGAGGG TCTCTTCGTG GGGTTTCCTG GTGTGCAGTG AGACGCACCA
175  C  F  E  E  N  I  Q  V  G  P  Y  P  S  A  G  L  S  F  I  E  W  L  L  V  G  F  S  W  V  D  S  Q  T  T

2501 GTACACCTTG TCGAAGATGC TTTCAGGGGC CATCCACTTC AGGGGCAGCC GGGCACTGCC CTTGCGGACG TAGTCGGGGT CTTTGTAGAT GTCCCGGGCA
     CATGTGGAAC AGCTTCTACG AAAGTCCCCG GTAGGTGAAG TCCCCGTCGG CCCGTGACGG GAACGCCTGC ATCAGCCCCA GAAACATCTA CAGGGCCCGT
141  Y  V  K  D  F  I  S  E  P  A  M  W  K  L  P  L  R  A  S  G  K  R  V  Y  D  P  D  K  Y  I  D  R  A

2601 AGGCCAAAGT CACAGATCTT CACCACGTCG CTTTCCGACA GCAGAATGTT CCGAGCAGCC AGGTCTCTGT GGATGCACTT TCGGGAAGCC AGGAACTCCA
     TCCGGTTTCA GTGTCTAGAA GTGGTGCAGC GAAAGGCTGT CGTCTTACAA GGCTCGTCGG TCCAGAGACA CCTACGTGAA AGCCCTTCGG TCCTTGAGGT
108  L  G  F  D  C  I  K  V  V  D  S  E  S  L  L  I  N  R  A  A  L  D  R  H  I  C  K  R  S  A  L  P  E

2701 TCCCTCTCGC CACCTGGAAG CTGTAGCAGA CAAGATCTTC CATGGTCAGC GGGCTCAGCC ACAGGTCCTC AGCTTCTTGG TCTGGAGAAG CCCGCCTCGC
     AGGGAGACCG GTGGACCTTC GACATCGTCT GTTCTAGAAG GTACCAGTCG CCCGAGTCGG TGTCCAGGAG TCGAAGAACC AGACCTCTTC GGGCGGAGCG
75   M  G  R  A  V  Q  P  S  Y  C  V  L  D  E  M  T  L  P  S  L  W  L  D  E  A  E  Q  D  P  S  A  R  R  A
```

FIGURE 4D

FIGURE 4E

```
2801 TCCGCCCTCG GTCTTCGAGA ACCGCGCGAA GAGGACCCTG TCGCTGCTCC CCGGCCGCCT CCGATCCAGC CTGGCGAGCT CCACCATGGC GCGGAAGCGT
     AGGCGGGAGC CAGAAGCTCT TGGCGCGCTT CTCCTGGGAC AGCGACGAGG GGCCGGCGGA GGCTAGGTCG GACCGCTCGA GGTGGTACCG CGCCTTCGCA
  41  C  G  E   T  K  S   F  R  A  F   L  V  R   D  S  S   G  P  R  R   R  D  L   R  A  L   E  V  M  A   R  F  R

2901 CCGCGCTGCT CGGGAGACTT CTCCTGCGGA TGCACGAAGC TGGCTCGAGG GCGCCCAGTC GTCCGCCGCA GAGGCGCCTC CATTCCCCCG CCGCCCGCGG
     GGCGCGACGA GCCCTCTGAA GAGGACGCCT ACGTGCTTCG ACCGAGCTCC CGCGGGTCAG CAGGCGGCGT CTCCGCGGAG GTAAGGGGGC GGCGGGCGCC
   8  G  R  Q   E  P  S  K   E

3001 CGCCCCGCAG GCCGCCCGCT CACCGXGCAG GGGCTGCGGC CGCGACTCTA GAGTCGACCT GCAGAAGCTT GGCCGCCATG GCCCAACTTG TTTATTGCAG
     GCGGGGCGTC CGGCGGGCGA GTGGC?CGTC CCCGACGCCG GCGCTGAGAT CTCAGCTGGA CGTCTTCGAA CCGGCGGTAC CGGGTTGAAC AAATAACGTC

3101 CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATCT
     GAATATTACC AATGTTTATT TCGTTATCGT AGTGTTTAAA GTGTTTATTT CGTAAAAAAA GTGACGTAAG ATCAACACCA AACAGGTTTG AGTAGTTACA

3201 ATCTTATCAT GTCTGGATCG ATCGGGAATT AATTCGGCGC AGCACCATGG CCTGAAATAA CCTCTGAAAG AGGAACTTGG TTAGGTACCT TCTGAGGCGG
     TAGAATAGTA CAGACCTAGC TAGCCCTTAA TTAAGCCGCG TCGTGGTACC GGACTTTATT GGAGACTTTC TCCTTGAACC AATCCATGGA AGACTCCGCC

3301 AAAGAACCAG CTGTGGAATG TGTGTCAGTT AGGGTGTGGA AAGTCCCCAG GCTCCCCAGC AGGCAGAAGT ATGCAAAGCA TGCATCTCAA TTAGTCAGCA
     TTTCTTGGTC GACACCTTAC ACACAGTCAA TCCCACACCT TTCAGGGGTC CGAGGGGTCG TCCGTCTTCA TACGTTTCGT ACGTAGAGTT AATCAGTCGT

3401 ACCAGGTGTG GAAAGTCCCC AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC AATTAGTCAG CAACCATAGT CCCGCCCCTA ACTCCGCCCA
     TGGTCCACAC CTTTCAGGGG TCCGAGGGGT CGTCCGTCTT CATACGTTTC GTACGTAGAG TTAATCAGTC GTTGGTATCA GGGCGGGGAT TGAGGCGGGT
```

FIGURE 4F

```
3501 TCCCGCCCCT AACTCCGCCC AGTTCCGCCC ATTCTCCGCC CCATGGCTGA CTAATTTTTT TTATTTATGC AGAGGCCGAG GCCGCCTCGG CCTCTGAGCT
     AGGGCGGGGA TTGAGGCGGG TCAAGGCGGG TAAGAGGCGG GGTACCGACT GATTAAAAAA AATAAATACG TCTCCGGCTC CGGCGGAGCC GGAGACTCGA

3601 ATTCCAGAAG TAGTGAGGAG GCTTTTTTGG AGGCCTAGGC TTTTGCAAAA AGCTGTTAAC AGCTTGGCAC TGGCCGTCGT TTTACAACGT CGTGACTGGG
     TAAGGTCTTC ATCACTCCTC CGAAAAAACC TCCGGATCCG AAAACGTTTT TCGACAATTG TCGAACCGTG ACCGGCAGCA AAATGTTGCA GCACTGACCC

3701 AAAACCCTGG CGTTACCCAA CTTAATCGCC TTGCAGCACA TCCCCCCTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC CTTCCCAACA
     TTTTGGGACC GCAATGGGTT GAATTAGCGG AACGTCGTGT AGGGGGGAAG CGGTCGACCG CATTATCGCT TCTCCGGGCG TGGCTAGCGG GAAGGGTTGT

3801 GTTGCGTAGC CTGAATGGCG AATGGCGCCT GATGCGGTAT TTTCTCCTTA CGCATCTGTG CGGTATTTCA CACCGCATAC GTCAAAGCAA CCATAGTACG
     CAACGCATCG GACTTACCGC TTACCGCGGA CTACGCCATA AAAGAGGAAT GCGTAGACAC GCCATAAAGT GTGGCGTATG CAGTTTCGTT GGTATCATGC

3901 CGCCCTGTAG CGGCGCATTA AGCGCGGCGG GTGTGGTGGT TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT TCGCTTTCTT
     GCGGGACATC GCCGCGTAAT TCGCGCCGCC CACACCACCA ATGCGCGTCG CACTGGCGAT GTGAACGGTC GCGGGATCGC GGGCGAGGAA AGCGAAAGAA

4001 CCCTTCCTTT CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAAATC GGGGGCTCCC TTTAGGGTTC CGATTTAGTG CTTTACGGCA CCTCGACCCC
     GGGAAGGAAA GAGCGGTGCA AGCGGCCGAA AGGGGCAGTT CGAGATTTAG CCCCCGAGGG AAATCCCAAG GCTAAATCAC GAAATGCCGT GGAGCTGGGG

4101 AAAAAACTTG ATTTGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA GACGGTTTTT CGCCCTTTGA CGTTGGAGTC CACGTTCTTT AATAGTGGAC
     TTTTTTGAAC TAAACCCACT ACCAAGTGCA TCACCCGGTA GCGGGACTAT CTGCCAAAAA GCGGGAAACT GCAACCTCAG GTGCAAGAAA TTATCACCTG
```

FIGURE 4G

```
4201 TCTTGTTCCA AACTGGAACA ACACTCAACC CTATCTCGGG CTATTCTTTT GATTTATAAG GGATTTTGCC GATTTCGGCC TATTGGTTAA AAAATGAGCT
     AGAACAAGGT TTGACCTTGT TGTGAGTTGG GATAGAGCCC GATAAGAAAA CTAAATATTC CCTAAAACGG CTAAAGCCGG ATAACCAATT TTTTACTCGA

4301 GATTTAACAA AAATTTAACG CGAATTTTAA CAAAATATTA ACGTTTACAA TTTTATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC ATAGTTAAGC
     CTAAATTGTT TTTAAATTGC GCTTAAAATT GTTTTATAAT TGCAAATGTT AAAATACCAC GTGAGAGTCA TGTTAGACGA GACTACGGCG TATCAATTCG

4401 CAACTCCGCT ATCGCTACGT GACTGGGTCA TGGCTGCGCC CCGACACCCG CCAACACCCG CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC
     GTTGAGGCGA TAGCGATGCA CTGACCCAGT ACCGACGCGG GGCTGTGGGC GGTTGTGGGC GACTGCGCGG GACTGCCCGA ACAGACGAGG GCCGTAGGCG

4501 TTACAGACAA GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGGCAGT ATTCTTGAAG ACGAAAGGGC
     AATGTCTGTT CGACACTGGC AGAGGCCCTC GACGTACACA GTCTCCAAAA GTGGCAGTAG TGGCTTTGCG CGCTCCGTCA TAAGAACTTC TGCTTTCCCG

4601 CTCGTGATAC GCCTATTTTT ATAGGTTAAT GTCATGATAA TAATGGTTTC TTAGACGTCA GGTGGCACTT TTCGGGGAAA TGTGCGCGGA ACCCCTATTT
     GAGCACTATG CGGATAAAAA TATCCAATTA CAGTACTATT ATTACCAAAG AATCTGCAGT CCACCGTGAA AAGCCCCTTT ACACGCGCCT TGGGGATAAA

4701 GTTTATTTTT CTAAATACAT TCAAATATGT ATCCGCTCAT GAGACAATAA CCCTGATAAA TGCTTCAATA ATATTGAAAA AGGAAGAGTA TGAGTATTCA
     CAAATAAAAA GATTTATGTA AGTTTATACA TAGGCGAGTA CTCTGTTATT GGGACTATTT ACGAAGTTAT TATAACTTTT TCCTTCTCAT ACTCATAAGT

4801 ACATTTCCGT GTCGCCCTTA TTCCCTTTTT TGCGGCATTT TGCCTTCCTG TTTTTGCTCA CCCAGAAACG CTGGTGAAAG TAAAAGATGC TGAAGATCAG
     TGTAAAGGCA CAGCGGGAAT AAGGGAAAAA ACGCCGTAAA ACGGAAGGAC AAAAACGAGT GGGTCTTTGC GACCACTTTC ATTTTCTACG ACTTCTAGTC
```

FIGURE 4H

```
4901 TTGGGTGCAC GAGTGGGTTA CATCGAACTG GATCTCAACA GCGGTAAGAT CCTTGAGAGT TTTCGCCCCG AAGAACGTTT TCCAATGATG AGCACTTTT
     AACCCACGTG CTCACCCAAT GTAGCTTGAC CTAGAGTTGT CGCCATTCTA GGAACTCTCA AAAGCGGGGC TTCTTGCAAA AGGTTACTAC TCGTGAAAA

5001 AAGTTCTGCT ATGTGGCGCG GTATTATCCC GTGATGACGC CGGGCAAGAG CAACTCGGTC GCCGCATACA CTATTCTCAG AATGACTTGG TTGAGTACTC
     TTCAAGACGA TACACCGCGC CATAATAGGG CACTACTGCG GCCCGTTCTC GTTGAGCCAG CGGCGTATGT GATAAGAGTC TTACTGAACC AACTCATGAG

5101 ACCAGTCACA GAAAAGCATC TTACGGATGG CATGACAGTA AGAGAATTAT GCAGTGCTGC CATAACCATG AGTGATAACA CTGCGGCCAA CTTACTTCTG
     TGGTCAGTGT CTTTTCGTAG AATGCCTACC GTACTGTCAT TCTCTTAATA CGTCACGACG GTATTGGTAC TCACTATTGT GACGCCGGTT GAATGAAGAC

5201 ACAACGATCG GAGGACCGAA GGAGCTAACC GCTTTTTTGC ACAACATGGG GGATCATGTA ACTCGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA
     TGTTGCTAGC CTCCTGGCTT CCTCGATTGG CGAAAAAACG TGTTGTACCC CCTAGTACAT TGAGCGGAAC TAGCAACCCT TGGCCTCGAC TTACTTCGGT

5301 TACCAAACGA CGAGCGTGAC ACCACGATGC CAGCAGCAAT GGCAACAACG TTGCGCAAAC TATTAACTGG CGAACTACTT ACTCTAGCTT CCCGGCAACA
     ATGGTTTGCT GCTCGCACTG TGGTGCTACG GTCGTCGTTA CCGTTGTTGC AACGCGTTTG ATAATTGACC GCTTGATGAA TGAGATCGAA GGGCCGTTGT

5401 ATTAATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA CTTCTGCGCT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG
     TAATTATCTG ACCTACCTCC GCCTATTTCA ACGTCCTGGT GAAGACGCGA GCCGGGAAGG CCGACCGACC AAATAACGAC TATTTAGACC TCGGCCACTC

5501 CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC
     GCACCCAGAG CGCCATAGTA ACGTCGTGAC CCCGGTCTAC CATTCGGGAG GGCATAGCAT CAATAGATGT GCTGCCCCTC AGTCCGTTGA TACCTACTTG
```

FIGURE 4I

```
5601 GAAATAGACA GATCGCTGAG ATAGGTGCCT CACTGATTAA GCATTGGTAA CTGTCAGACC AAGTTTACTC ATATATACTT TAGATTGATT TAAAACTTC
     CTTTATCTGT CTAGCGACTC TATCCACGGA GTGACTAATT CGTAACCATT GACAGTCTGG TTCAAATGAG TATATATGAA ATCTAACTAA ATTTTGAAG

5701 TTTTTAATTT AAAAGGATCT AGGTGAAGAT CCTTTTTGAT AATCTCATGA CCAAAATCCC TTAACGTGAG TTTTCGTTCC ACTGAGCGTC AGACCCCGTA
     AAAAATTAAA TTTTCCTAGA TCCACTTCTA GGAAAAACTA TTAGAGTACT GGTTTTAGGG AATTGCACTC AAAAGCAAGG TGACTCGCAG TCTGGGGCAT

5801 GAAAAGATCA AAGGATCTTC TTGAGATCCT TTTTTTCTGC GCGTAATCTG CTGCTTGCAA ACAAAAAAAC CACCGCTACC AGCGGTGGTT TGTTTGCCGG
     CTTTTCTAGT TTCCTAGAAG AACTCTAGGA AAAAAAGACG CGCATTAGAC GACGAACGTT TGTTTTTTTG GTGGCGATGG TCGCCACCAA ACAAACGGCC

5901 ATCAAGAGCT ACCAACTCTT TTTCCGAAGG TAACTGGCTT CAGCAGAGCG CAGATACCAA ATACTGTCCT TCTAGTGTAG CCGTAGTTAG GCCACCACTT
     TAGTTCTCGA TGGTTGAGAA AAAGGCTTCC ATTGACCGAA GTCGTCTCGC GTCTATGGTT TATGACAGGA AGATCACATC GGCATCAATC CGGTGGTGAA

6001 CAAGAACTCT GTAGCACCGC CTACATACCT CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC TGCCAGTGGC GATAAGTCGT GTCTTACCGG GTTGGACTCA
     GTTCTTGAGA CATCGTGGCG GATGTATGGA GCGAGACGAT TAGGACAATG GTCACCGACG ACGGTCACCG CTATTCAGCA CAGAATGGCC CAACCTGAGT

6101 AGACGATAGT TACCGGATAA GGCGCAGCGG TCGGGCTGAA CGGGGGGTTC GTGCACACAG CCCAGCTTGG AGCGAACGAC CTACACCGAA CTGAGATACC
     TCTGCTATCA ATGGCCTATT CCGCGTCGCC AGCCCGACTT GCCCCCCAAG CACGTGTGTC GGGTCGAACC TCGCTTGCTG GATGTGGCTT GACTCTATGG

6201 TACAGCGTGA GCATTGAGAA AGCGCCACGC TTCCCGAAGG GAGAAAGGCG GACAGGTATC CGGTAAGCGG CAGGGTCGGA ACAGGAGAGC GCACGAGGGA
     ATGTCGCACT CGTAACTCTT TCGCGGTGCG AAGGGCTTCC CTCTTTCCGC CTGTCCATAG GCCATTCGCC GTCCCAGCCT TGTCCTCTCG CGTGCTCCCT
```

FIGURE 4J

```
6301 GCTTCCAGGG GGAAACGCCT GGTATCTTTA TAGTCCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA TTTTTGTGAT GCTCGTCAGG GGGGCGGAGC
     CGAAGGTCCC CCTTTGCGGA CCATAGAAAT ATCAGGACAG CCCAAAGCGG TGGAGACTGA ACTCGCAGCT AAAAACACTA CGAGCAGTCC CCCCGCCTCG

6401 CTATGGAAAA ACGCCAGCAA CGCGGCCTTT TTACGGTTCC TGGCCTTTTG CTGGCCTTTT GCTCACATGT TCTTTCCTGC GTTATCCCCT GATTCTGTGG
     GATACCTTTT TGCGGTCGTT GCGCCGGAAA AATGCCAAGG ACCGGAAAAC GACCGGAAAA CGAGTGTACA AGAAAGGACG CAATAGGGGA CTAAGACACC

6501 ATAACCGTAT TACCGCCTTT GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT
     TATTGGCATA ATGGCGGAAA CTCACTCGAC TATGGCGAGC GGCGTCGGCT TGCTGGCTCG CGTCGCTCAG TCACTCGCTC CTTCGCCTTC TCGCGGGTTA

6601 ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TCCAGCTGGC ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT
     TGCGTTTGGC GGAGAGGGGC GCGCAACCGG CTAAGTAATT AGGTCGACCG TGCTGTCCAA AGGGCTGACC TTTCGCCCGT CACTCGCGTT GCGTTAATTA

6701 GTGAGTTACC TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA TTGTGAGCGG ATAACAATTT CACACAGGAA
     CACTCAATGG AGTGAGTAAT CCGTGGGGTC CGAAATGTGA AATACGAAGG CCGAGCATAC AACACACCTT AACACTCGCC TATTGTTAAA GTGTGTCCTT

6801 ACAGCTATGA CCATGATTAC GAATTAA
     TGTCGATACT GGTACTAATG CTTAATT
```

FIGURE 5A

```
  1 TTCGAGCTCG CCCGACATTG ATTATTGACT AGTTATTAAT AGTAATCAAT TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC
    AAGCTCGAGC GGGCTGTAAC TAATAACTGA TCAATAATTA TCATTAGTTA ATGCCCCAGT AATCAAGTAT CGGGTATATA CCTCAAGGCG CAATGTATTG

101 TTACGGTAAA TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA
    AATGCCATTT ACCGGGCGGA CCGACTGGCG GGTTGCTGGG GGCGGGTAAC TGCAGTTATT ACTGCATACA AGGGTATCAT TGCGGTTATC CCTGAAAGGT

201 TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTACGCCCC CTATTGACGT CAATGACGGT
    AACTGCAGTT ACCCACCTCA TAAATGCCAT TTGACGGGTG AACCGTCATG TAGTTCACAT AGTATACGGT TCATGCGGGG GATAACTGCA GTTACTGCCA

301 AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC
    TTTACCGGGC GGACCGTAAT ACGGGTCATG TACTGGAATA CCCTGAAAGG ATGAACCGTC ATGTAGATGC ATAATCAGTA GCGATAATGG TACCACTACG

401 GGTTTTGGCA GTACATCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA
    CCAAAACCGT CATGTAGTTA CCCGCACCTA TCGCCAAACT GAGTGCCCCT AAAGGTTCAG AGGTGGGGTA ACTGCAGTTA CCCTCAAACA AAACCGTGGT

501 AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC CCATTGACGC AAATGGGCGG TAGGCGTGTA CGGTGGGAGG TCTATATAAG CAGAGCTCGT
    TTTAGTTGCC CTGAAAGGTT TTACAGCATT GTTGAGGCGG GGTAACTGCG TTTACCCGCC ATCCGCACAT GCCACCCTCC AGATATATTC GTCTCGAGCA

601 TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT CCATAGAAGA CACCGGGACC GATCCAGCCT CCGCGGCCGG GAACGGTGCA
    AATCACTTGG CAGTCTAGCG GACCTCTGCG GTAGGTGCGA CAAAACTGGA GGTATCTTCT GTGGCCCTGG CTAGGTCGGA GGCGCCGGCC CTTGCCACGT

701 TTGGAACGCG GATTCCCCGT GCCAAGAGTG ACGTAAGTAC CGCCTATAGA GTCTATAGGC CCACTTGGCT TCGTTAGAAC GCGGCTACAA TTAATACATA
    AACCTTGCGC CTAAGGGGCA CGGTTCTCAC TGCATTCATG GCGGATATCT CAGATATCCG GGTGAACCGA AGCAATCTTG CGCCGATGTT AATTATGTAT
```

FIGURE 5B

```
801  ACCTTATGTA TCATACACAT ACGATTTAGG TGACACTATA GAATAACATC CACTTTGCCT TTCTCTCCAC AGGTGTCCAC TCCCAGGTCC AACTGCACCT
     TGGAATACAT AGTATGTGTA TGCTAAATCC ACTGTGATAT CTTATTGTAG GTGAAACGGA AAGAGAGGTG TCCACAGGTG AGGGTCCAGG TTGACGTGGA

901  CGGTTCTATC GATTGAATTC CCCGGGGATC CTCTAGAGAT CCCTCGACCT CGAGTCGACT TTTTTTTTTT TTTTTGTAGG CCAAAGGGTA CTTCTTTTTC
     GCCAAGATAG CTAACTTAAG GGGCCCCTAG GAGATCTCTA GGGAGCTGGA GCTCAGCTGA AAAAAAAAAA AAAAACATCC GGTTTCCCAT GAAGAAAAAG

1001 TTTATTAATT ACTCAGAAGT CTAGGCCACA GCAATCTACT GTTCTCCTCT CATTTTCCTA AACTATTTTG ATACCTATTT CTCAGACTTT ATGGGCTATT
     AAATAATTAA TGAGTCTTCA GATCCGGTGT CGTTAGATGA CAAGAGGAGA GTAAAAGGAT TTGATAAAAC TATGGATAAA GAGTCTGAAA TACCCGATAA

1101 AGACATTTCT CACATTTCCA TAGATAATAA CTCATCCGTT TTGCAACCTG ATTCTCAATA TTAAGAGATT AAAACTAATG TATATGACTC TCAGTTGACA
     TCTGTAAAGA GTGTAAAGGT ATCTATTATT GAGTAGGCAA AACGTTGGAC TAAGAGTTAT AATTCTCTAA TTTTGATTAC ATATACTGAG AGTCAACTGT

1201 CATACTGAAG TACAGAAAAA TTCCATCATT TCCTTCTGCA AAATGAAAAA GACTTCGTTT TCTCAACAGC TGCATCATTT TTTTATGCAT AGAAAAAAAT
     GTATGACTTC ATGTCTTTTT AAGGTAGTAA AGGAAGACGT TTTACTTTTT CTGAAGCAAA AGAGTTGTCG ACGTAGTAAA AAAATACGTA TCTTTTTTTA

1301 GTGCAATTAC TCCAAGTACA ATCAAGTCAT TTAACATGGC TTTACCATCA TTGTAGTTAC AGGATATTTT AAAAGAGAAA AAAAAATCTC AAAGCACAGG
     CACGTTAATG AGGTTCATGT TAGTTCAGTA AATTGTACCG AAATGGTAGT AACATCAATG TCCTATAAAA TTTTCTCTTT TTTTTTAGAG TTTCGTGTCC

1401 TCCTGCTGTG CAGCAAAGCA ATCAAATTCC TTCATAATAA CAGCCTGATG GGATTCAGCA ATCTGAGGAA TAATGAATAA CCACTCTAAT CAGTAAACAG
     AGGACGACAC GTCGTTTCGT TAGTTTAAGG AAGTATTATT GTCGGACTAC CCTAAGTCGT TAGACTCCTT ATTACTTATT GGTGAGATTA GTCATTTGTC

1501 GAAAATGCTA CAACAGTCAC TGAGTAAAAA TTGGACTATC ATCTGTTGAT TCTCTTGATC GACATTTCAA ACAATAAATG GAAATGTAAG TATCTCTTAA
     CTTTTACGAT GTTGTCAGTG ACTCATTTTT AACCTGATAG TAGACAACTA AGAGAACTAG CTGTAAAGTT TGTTATTTAC CTTTACATTC ATAGAGAATT
```

FIGURE 5C

```
1601 AAAGAAAAAT AACTTGGTTT AGTGTGCTTA ATTTTACCAG GCAGTGAGGA AATTATATAT CACCTTGACT GTCCTGCAGT GTTGCCCAGT CAATAAAATG
     TTTCTTTTTA TTGAACCAAA TCACACGAAT TAAAATGGTC CGTCACTCCT TTAATATATA GTGGAACTGA CAGGACGTCA CAACGGGTCA GTTATTTTAC

1701 CACAAATAAT CTTTTTCATA ATACATGGCC AACTTTATCC TATCACTTGA ATATGTCAGG ATAAACTGAT TGTGCAGTTG GTTGATAACA TTGTATTTTG
     GTGTTTATTA GAAAAAGTAT TATGTACCGG TTGAAATAGG ATAGTGAACT TATACAGTCC TATTTGACTA ACACGTCAAC CAACTATTGT AACATAAAAC

1801 GAATGGATTA TTTGAATTTG TTTTGCTACT TTATTATTTG ATATTCTTCT CCAGTGTTCA TCTTATGAAG TTATTTGCAT CTGAATATGA AGAGTCTGTT
     CTTACCTAAT AAACTTAAAC AAAACGATGA AATAATAAAC TATAAGAAGA GGTCACAAGT AGAATACTTC AATAAACGTA GACTTATACT TCTCAGACAA
 506                                                             Q  R  I  P   N  N  A   D  S  Y  S   S  D  T

1901 TCAAAATAGT CTTCAAGTTT CCAACGCAGT GTCTCAAATG TAGGTCGTTC CTTAGGCTCT GCATTCCAGC ACTCCAACAT GATGTTGTAA AATTGCTGTG
     AGTTTTATCA GAAGTTCAAA GGTTGCGTCA CAGAGTTTAC ATCCAGCAAG GAATCCGAGA CGTAAGGTCG TGAGGTTGTA CTACAACATT TTAACGACAC
 492 E  F  Y   D  E  L  K  W  R  L   T  E  F   T  P  R  E   K  P  E   A  N  W   C  E  L  M   I  N  Y   F  Q  Q

2001 GACAGTTGGA TGGTTGCGGA AGTCTATAGT TTTGAGCCAA CATCTGGATT ACCTGGGCAC CTGTCATACC ACTGTAAGGC ATTTTGCCAT AAGTAATGAT
     CTGTCAACCT ACCAACGCCT TCAGATATCA AAACTCGGTT GTAGACCTAA TGGACCCGTG GACAGTATGG TGACATTCCG TAAAACGGTA TTCATTACTA
 459 P  C  N  S   P  Q  P   L  R  Y   N  Q  A  L   M  Q  I   V  Q  A   G  T  M  G   S  Y  P   M  K  G   Y  T  I  I

2101 TTCATAAAGA AGGATTCCAA ATGACCATAC ATCGGACTTA ATGCTGAATT TATTACTACG AATGGCTTCG GGCGCAGTCC ACTTCACCGG CAGCTTTATT
     AAGTATTTCT TCCTAAGGTT TACTGGTATG TAGCCTGAAT TACGACTTAA ATAATGATGC TTACCGAAGC CCGCGTCAGG TGAAGTGGCC GTCGAAATAA
 425  E  Y  L   L  I  G   F  S  W  V   D  S  K   I  S  F   K  N  S  R   I  A  E   P  A  T   W  K  V  P   L  K  I

2201 TCGTGTCTAG ATTCATAGAT GTCTTCATTA TCTACCTTAA AAACTCTGGC AAGTCCAAAA TCTGCTACTT TGTAGATATT ATGTTCACCA ACGAGGACAT
     AGCACAGATC TAAGTATCTA CAGAAGTAAT AGATGGAATT TTTGAGACCG TTCAGGTTTT AGACGATGAA ACATCTATAA TACAAGTGGT TGCTCCTGTA
 392  E  H  R   S  E  Y  I   D  E  N   D  V  K   F  V  R  A   L  G  F   D  A  V   K  Y  I  N   H  E  G   V  L  V

2301 TTCTGGCAGC CAGATCTCTG TGAATGTAGT TCCGAGACTC CAGATAGGCC ATTCCAGAGG CAACCTGTGC CGCCATGTCT ACCTGTTGAG TCAGATGGAT
     AAGACCGTCG GTCTAGAGAC ACTTACATCA AGGCTCTGAG GTCTATCCGG TAAGGTCTCC GTTGGACACG GCGGTACAGA TGGACAACTC AGTCTACCTA
 359 N  R  A  A   L  D  R   H  I  Y   N  R  S  E   L  Y  A   M  G  S   A  V  Q  A   A  M  D   V  Q  Q   T  L  H  I
```

FIGURE 5D

```
2401 TTTTGATCCA GTGTCATTTT GGAGATATTC TTGCAGACTT CCATGTCTCA TCAACTCTGT AATAATATAA ATTGGATCTT CTAAAGTGCA AACAGCATAA
     AAAACTAGGT CACAGTAAAA CCTCTATAAG AACGTCTGAA GGTACAGAGT AGTTGAGACA TTATTATATT TAACCTAGAA GATTTCACGT TTGTCGTATT
 325  K  S  G   T  D  N   Q  L  Y  E   Q  L  S   G  H  R   M  L  E  T   I  I  Y   I  P  D   E  L  T  C   V  A  Y

2501 AGCTGGATAA GCTTTGGATG TCTTAGGTTC TTCATTATCT GTGCCTCCCT CAGGAAGTCA TTTGGATCCA TTGAACCTGG TTTTAATGTT TTCACTGCTA
     TCGACCTATT CGAAACCTAC AGAATCCAAG AAGTAATAGA CACGGAGGGA GTCCTTCAGT AAACCTAGGT AACTTGGACC AAAATTACAA AAGTGACGAT
 292 L  Q  I   L  K  P  H   R  L  N   K  M  I   Q  A  E  R   L  F  D   N  P  D   M  S  G  P   K  L  T   K  V  A

2601 CTGGAGTGGT ATTGTTCCAC AGACCTTCCC ATACTTCGCC AAACTGACCA GATCCCAATC GCTTCAGAAG CTGTATGGAG TTGCGGTCTA TCTCCCATTG
     GACCTCACCA TAACAAGGTG TCTGGAAGGG TATGAAGCGG TTTGACTGGT CTAGGGTTAG CGAAGTCTTC GACATACCTC AACGCCAGAT AGAGGGTAAC
 259 V  P  T  T   N  N  W   L  G  E   W  V  E  G   F  Q  G   S  G  L   R  K  L  L   Q  I  S   N  R  D   I  E  W  Q

2701 GTCCACGGTT TTATACGACA AATCAAATGG AGCTGGGACC TGGATCTTTA AGCATGGTTT CCCCAGCTTG ACACACAGGC CGTCACTTGT CTTGGTGTAG
     CAGGTGCCAA AATATGCTGT TTAGTTTACC TCGACCCTGG ACCTAGAAAT TCGTACCAAA GGGGTCGAAC TGTGTGTCCG GCAGTGAACA GAACCACATC
 225  D  V  T   K  Y  S   L  D  F  P   A  P  V   Q  I  K   L  C  P  K   G  L  K   V  C  L   G  D  S  T   K  T  Y

2801 TGGCTCACAA ATTCGTTCAG TGTTGAAAAG ATTCTTCTTC GCGTGAGAAA AAATCCCCCT TCATCCAGTC TTTTAATTCT GTAGTGTTTT ACAACTGCTC
     ACCGAGTGTT TAAGCAAGTC ACAACTTTTC TAAGAAGAAG CGCACTCTTT TTTAGGGGGA AGTAGGTCAG AAAATTAAGA CATCACAAAA TGTTGACGAG
 192 H  S  V   F  E  N  L   T  S  F   I  R  R   R  T  L  F   F  G  G   E  D  L   R  K  I  R   Y  H  K   V  V  A

2901 CATCTAAAAC TGAAAGAGAG AATTCTCCTT TTTGGCTTTC ACTTTCTCTG ATTAGAAAGG AACCGGTCTT GTTTTCTGAA TATAATAGTT GTTTCTCTGC
     GTAGATTTTG ACTTTCTCTC TTAAGAGGAA AAACCGAAAG TGAAAGAGAC TAATCTTTCC TTGGCCAGAA CAAAAGACTT ATATTATCAA CAAAGAGACG
 159 G  D  L  V   S  L  S   F  E  G   K  Q  S  E   S  E  R   I  L  F   S  G  T  K   N  E  S   Y  L  L   Q  K  E  A

3001 ATCTGATCTT CCGATTGCTC CAAAGAACCA CGGCTCTGCC TGTAGGCTTC TGTCCTCAGC CACGTAGTTA GAAGGAATAT AGCCTTGTAG TTGCTGACTG
     TAGACTAGAA GGCTAACGAG GTTTCTTGGT GCCGAGACGG ACATCCGAAG ACAGGAGTCG GTGCATCAAT CTTCCTTATA TCGGAACATC AACGACTGAC
 125  D  S  R   G  I  A   G  F  F  W   P  E  A   Q  L  S   R  D  E  A   V  Y  N   S  P  I   Y  G  Q  L   Q  Q  S

3101 GAGCCATCTC GTCTTTTCTC CAAGTGTCTG GCAAACCACC AGCCCTCATG CAAAGTGTCC AGAACTTGAA GTTTGTCACC TGCTCGGAAG CTCAAGTCCT
     CTCGGTAGAG CAGAAAAGAG GTTCACAGAC CGTTTGGTGG TCGGGAGTAC GTTTCACAGG TCTTGAACTT CAAACAGTGG ACGAGCCTTC GAGTTCAGGA
  92 S  G  D   R  R  K  E   L  H  R   A  F  W   W  G  E  H   L  T  D   L  V  Q   L  K  D  G   A  R  F   S  L  D
```

FIGURE 5E

```
3201 CAGCAGTCCG AGCCTGGTAA TCAAACAAAG CCACAAAGTA GTGGCCATGC CTCTGTGACT GGGGAGAGCA AAGGGCCCCT GGATTTTCAA TCACGGTTGA
     GTCGTCAGGC TCGGACCATT AGTTTGTTTC GGTGTTTCAT CACCGGTACG GAGACACTGA CCCCTCTCGT TTCCCGGGGA CCTAAAAGTT AGTGCCAACT
  59 E  A  T  R   A  Q  Y   D  F  L   A  V  F  Y   H  G  H   R  Q  S   Q  P  S  C   L  A  G   P  N  E   I  V  T  S

3301 CTTGTCTGCC TCCGTGGACA AACAGGGGAG ATAGGGTTCT AGGTACTCCC AGAGCCTCTG ACAGATGTTG CTCATTGTGC CTTGGTGGGG AGAAGAGGAG
     GAACAGACGG AGGCACCTGT TTGTCCCCTC TATCCCAAGA TCCATGAGGG TCTCGGAGAC TGTCTACAAC GAGTAACACG GAACCACCCC TCTTCTCCTC
  25  K  D  A   E  T  S   L  C  P  L   Y  P  E   L  Y  E   W  L  R  Q   C  I  N   S  M

3401 CAGGGCTTCT CCCTCTCCCC TTAGTCTCTG CGATCCACCT TATCTTCCTT CACCAGGCAA CTTTGAAGTC AGCACCAACT CACCATACTT CGGAGAGTAT
     GTCCCGAAGA GGGAGAGGGG AATCAGAGAC GCTAGGTGGA ATAGAAGGAA GTGGTCCGTT GAAACTTCAG TCGTGGTTGA GTGGTATGAA GCCTCTCATA

3501 GCAAAGTCCC GTTTCAGATC AGTCCAGCAG CTGGGTTGCA GCAAGTCCTA CCTGGAGAGA CTTACCGGCT TGCTTTCTGT GGCTGGAGGT GCTACCCCGA
     CGTTTCAGGG CAAAGTCTAG TCAGGTCGTC GACCCAACGT CGTTCAGGAT GGACCTCTCT GAATGGCCGA ACGAAAGACA CCGACCTCCA CGATGGGGCT

3601 GGCAAAACTG AGCAGGAGCT GGGCAGCTGC TCACTAGGAA GGTGTCTTTT CTTCTTATCT GCTTAAGAAT CCCACAACAA AAATAAAATA AAATTAAAAG
     CCGTTTTGAC TCGTCCTCGA CCCGTCGACG AGTGATCCTT CCACAGAAAA GAAGAATAGA CGAATTCTTA GGGTGTTGTT TTTATTTTAT TTTAATTTTC

3701 GGCTTTATTT AGACAAATAT CTGAGAACAG AATGGTGCCA TCTTGCCTTT TGTCCCAATA AAAAGTTAGC AAGAGGAAGC TACTAACCCC TGGTAAAACC
     CCGAAATAAA TCTGTTTATA GACTCTTGTC TTACCACGGT AGAACGGAAA ACAGGGTTAT TTTTCAATCG TTCTCCTTCG ATGATTGGGG ACCATTTTGG

3801 TCCACGTCTT GCTTTCGCCA GGGTCGACTC GAGGGATCTT CCATACCTAC CAGTTCTGCG CCTGCAGGTC GCGGCCGCGA CTCTAGAGTC GACCTGCAGA
     AGGTGCAGAA CGAAAGCGGT CCCAGCTGAG CTCCCTAGAA GGTATGGATG GTCAAGACGC GGACGTCCAG CGCCGGCGCT GAGATCTCAG CTGGACGTCT

3901 AGCTTGGCCG CCATGGCCCA ACTTGTTTAT TGCAGCTTAT AATGGTTACA AATAAAGCAA TAGCATCACA AATTTCACAA ATAAAGCATT TTTTTCACTG
     TCGAACCGGC GGTACCGGGT TGAACAAATA ACGTCGAATA TTACCAATGT TTATTTCGTT ATCGTAGTGT TTAAAGTGTT TATTTCGTAA AAAAAGTGAC
```

FIGURE 5F

```
4001 CATTCTAGTT GTGGTTTGTC CAAACTCATC AATGTATCTT ATCATGTCTG GATCGGGAAT TAATTCGGCG CAGCACCATG GCCTGAAATA ACCTCTGAAA
     GTAAGATCAA CACCAAACAG GTTTGAGTAG TTACATAGAA TAGTACAGAC CTAGCCCTTA ATTAAGCCGC GTCGTGGTAC CGGACTTTAT TGGAGACTTT

4101 GAGGAACTTG GTTAGGTACC TTCTGAGGCG GAAAGAACCA GCTGTGGAAT GTGTGTCAGT TAGGGTGTGG AAAGTCCCCA GGCTCCCCAG CAGGCAGAAG
     CTCCTTGAAC CAATCCATGG AAGACTCCGC CTTTCTTGGT CGACACCTTA CACACAGTCA ATCCCACACC TTTCAGGGGT CCGAGGGGTC GTCCGTCTTC

4201 TATGCAAAGC ATGCATCTCA ATTAGTCAGC AACCAGGTGT GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA
     ATACGTTTCG TACGTAGAGT TAATCAGTCG TTGGTCCACA CCTTTCAGGG GTCCGAGGGG TCGTCCGTCT TCATACGTTT CGTACGTAGA GTTAATCAGT

4301 GCAACCATAG TCCCGCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC CAGTTCCGCC CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG
     CGTTGGTATC AGGGCGGGGA TTGAGGCGGG TAGGGCGGGG ATTGAGGCGG GTCAAGGCGG GTAAGAGGCG GGGTACCGAC TGATTAAAAA AAATAAATAC

4401 CAGAGGCCGA GGCCGCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG CTTTTGCAAA AAGCTGTTAA CAGCTTGGCA
     GTCTCCGGCT CCGGCGGAGC CGGAGACTCG ATAAGGTCTT CATCACTCCT CCGAAAAAAC CTCCGGATCC GAAAACGTTT TTCGACAATT GTCGAACCGT

4501 CTGGCCGTCG TTTTACAACG TCGTGACTGG GAAAACCCTG GCGTTACCCA ACTTAATCGC CTTGCAGCAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG
     GACCGGCAGC AAAATGTTGC AGCACTGACC CTTTTGGGAC CGCAATGGGT TGAATTAGCG GAACGTCGTG TAGGGGGAAA GCGGTCGACC GCATTATCGC

4601 AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGAATGGC GAATGGCGCC TGATGCGGTA TTTTCTCCTT ACGCATCTGT GCGGTATTTC
     TTCTCCGGGC GTGGCTAGCG GGAAGGGTTG TCAACGCGTC GGACTTACCG CTTACCGCGG ACTACGCCAT AAAAGAGGAA TGCGTAGACA CGCCATAAAG

4701 ACACCGCATA CGTCAAAGCA ACCATAGTAC GCGCCCTGTA GCGGCGCATT AAGCGCGGCG GGTGTGGTGG TTACGCGCAG CGTGACCGCT ACACTTGCCA
     TGTGGCGTAT GCAGTTTCGT TGGTATCATG CGCGGGACAT CGCCGCGTAA TTCGCGCCGC CCACACCACC AATGCGCGTC GCACTGGCGA TGTGAACGGT
```

FIGURE 5G

```
4801 GCGCCCTAGC GCCCGCTCCT TTCGCTTTCT TCCCTTCCTT TCTCGCCACG TTCGCCGGCT TTCCCCGTCA AGCTCTAAAT CGGGGGCTCC CTTTAGGGTT
     CGCGGGATCG CGGGCGAGGA AAGCGAAAGA AGGGAAGGAA AGAGCGGTGC AAGCGGCCGA AAGGGGCAGT TCGAGATTTA GCCCCCGAGG GAAATCCCAA

4901 CCGATTTAGT GCTTTACGGC ACCTCGACCC CAAAAAACTT GATTTGGGTG ATGGTTCACG TAGTGGGCCA TCGCCCTGAT AGACGGTTTT TCGCCCTTTG
     GGCTAAATCA CGAAATGCCG TGGAGCTGGG GTTTTTTGAA CTAAACCCAC TACCAAGTGC ATCACCCGGT AGCGGGACTA TCTGCCAAAA AGCGGGAAAC

5001 ACGTTGGAGT CCACGTTCTT TAATAGTGGA CTCTTGTTCC AAACTGGAAC AACACTCAAC CCTATCTCGG GCTATTCTTT TGATTTATAA GGGATTTTGC
     TGCAACCTCA GGTGCAAGAA ATTATCACCT GAGAACAAGG TTTGACCTTG TTGTGAGTTG GGATAGAGCC CGATAAGAAA ACTAAATATT CCCTAAAACG

5101 CGATTTCGGC CTATTGGTTA AAAAATGAGC TGATTTAACA AAAATTTAAC GCGAATTTTA ACAAAATATT AACGTTTACA ATTTTATGGT GCACTCTCAG
     GCTAAAGCCG GATAACCAAT TTTTTACTCG ACTAAATTGT TTTTAAATTG CGCTTAAAAT TGTTTTATAA TTGCAAATGT TAAAATACCA CGTGAGAGTC

5201 TACAATCTGC TCTGATGCCG CATAGTTAAG CCAGCCCCGA CACCCGCCAA CACCCGCTGA CGCGCCCTGA CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC
     ATGTTAGACG AGACTACGGC GTATCAATTC GGTCGGGGCT GTGGGCGGTT GTGGGCGACT GCGCGGGACT GCCCGAACAG ACGAGGGCCG TAGGCGAATG

5301 AGACAAGCTG TGACCGTCTC CGGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG AAACGCGCGA GACGAAAGGG CCTCGTGATA CGCCTATTTT
     TCTGTTCGAC ACTGGCAGAG GCCCTCGACG TACACAGTCT CCAAAAGTGG CAGTAGTGGC TTTGCGCGCT CTGCTTTCCC GGAGCACTAT GCGGATAAAA

5401 TATAGGTTAA TGTCATGATA ATAATGGTTT CTTAGACGTC AGGTGGCACT TTTCGGGGAA ATGTGCGCGG AACCCCTATT TGTTTATTTT TCTAAATACA
     ATATCCAATT ACAGTACTAT TATTACCAAA GAATCTGCAG TCCACCGTGA AAAGCCCCTT TACACGCGCC TTGGGGATAA ACAAATAAAA AGATTTATGT

5501 TTCAAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA ATGCTTCAAT AATATTGAAA AAGGAAGAGT ATGAGTATTC AACATTTCCG TGTCGCCCTT
     AAGTTTATAC ATAGGCGAGT ACTCTGTTAT TGGGACTATT TACGAAGTTA TTATAACTTT TTCCTTCTCA TACTCATAAG TTGTAAAGGC ACAGCGGGAA
```

FIGURE 5H

5601 ATTCCCTTTT TTGCGGCATT TTGCCTTCCT GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA CGAGTGGGTT
TAAGGGAAAA AACGCCGTAA AACGGAAGGA CAAAAACGAG TGGGTCTTTG CGACCACTTT CATTTTCTAC GACTTCTAGT CAACCCACGT GCTCACCCAA

5701 ACATCGAACT GGATCTCAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC
TGTAGCTTGA CCTAGAGTTG TCGCCATTCT AGGAACTCTC AAAAGCGGGG CTTCTTGCAA AAGGTTACTA CTCGTGAAAA TTTCAAGACG ATACACCGCG

5801 GGTATTATCC CGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC AGAAAAGCAT
CCATAATAGG GCATAACTGC GGCCCGTTCT CGTTGAGCCA GCGGCGTATG TGATAAGAGT CTTACTGAAC CAACTCATGA GTGGTCAGTG TCTTTTCGTA

5901 CTTACGGATG GCATGACAGT AAGAGAATTA TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC GGAGGACCGA
GAATGCCTAC CGTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGGTA CTCACTATTG TGACGCCGGT TGAATGAAGA CTGTTGCTAG CCTCCTGGCT

6001 AGGAGCTAAC CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA
TCCTCGATTG GCGAAAAAAC GTGTTGTACC CCCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCTCGA CTTACTTCGG TATGGTTTGC TGCTCGCACT

6101 CACCACGATG CCTGTAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG GCGAACTACT TACTCTAGCT TCCCGGCAAC AATTAATAGA CTGGATGGAG
GTGGTGCTAC GGACATCGTT ACCGTTGTTG CAACGCGTTT GATAATTGAC CGCTTGATGA ATGAGATCGA AGGGCCGTTG TTAATTATCT GACCTACCTC

6201 GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG GTTTATTGCT GATAAATCTG GAGCCGGTGA GCGTGGGTCT CGCGGTATCA
CGCCTATTTC AACGTCCTGG TGAAGACGCG AGCCGGGAAG GCCGACCGAC CAAATAACGA CTATTTAGAC CTCGGCCACT CGCACCCAGA GCGCCATAGT

6301 TTGCAGCACT GGGGCCAGAT GGTAAGCCCT CCCGTATCGT AGTTATCTAC ACGACGGGGA GTCAGGCAAC TATGGATGAA CGAAATAGAC AGATCGCTGA
AACGTCGTGA CCCCGGTCTA CCATTCGGGA GGGCATAGCA TCAATAGATG TGCTGCCCCT CAGTCCGTTG ATACCTACTT GCTTTATCTG TCTAGCGACT

FIGURE 51

```
6401 GATAGGTGCC TCACTGATTA AGCATTGGTA ACTGTCAGAC CAAGTTTACT CATATATACT TTAGATTGAT TTAAAACTTC ATTTTTAATT TAAAAGGATC
     CTATCCACGG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAAATGA GTATATATGA AATCTAACTA AATTTTGAAG TAAAAATTAA ATTTTCCTAG

6501 TAGGTGAAGA TCCTTTTTGA TAATCTCATG ACCAAAATCC CTTAACGTGA GTTTTCGTTC CACTGAGCGT CAGACCCCGT AGAAAAGATC AAAGGATCTT
     ATCCACTTCT AGGAAAAACT ATTAGAGTAC TGGTTTTAGG GAATTGCACT CAAAAGCAAG GTGACTCGCA GTCTGGGGCA TCTTTTCTAG TTTCCTAGAA

6601 CTTGAGATCC TTTTTTTCTG CGCGTAATCT GCTGCTTGCA AACAAAAAAA CCACCGCTAC CAGCGGTGGT TTGTTTGCCG GATCAAGAGC TACCAACTCT
     GAACTCTAGG AAAAAAAGAC GCGCATTAGA CGACGAACGT TTGTTTTTTT GGTGGCGATG GTCGCCACCA AACAAACGGC CTAGTTCTCG ATGGTTGAGA

6701 TTTTCCGAAG GTAACTGGCT TCAGCAGAGC GCAGATACCA AATACTGTTC TTCTAGTGTA GCCGTAGTTA GGCCACCACT TCAAGAACTC TGTAGCACCG
     AAAAGGCTTC CATTGACCGA AGTCGTCTCG CGTCTATGGT TTATGACAAG AAGATCACAT CGGCATCAAT CCGGTGGTGA AGTTCTTGAG ACATCGTGGC

6801 CCTACATACC TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCG TGTCTTACCG GGTTGGACTC AAGACGATAG TTACCGGATA
     GGATGTATGG AGCGAGACGA TTAGGACAAT GGTCACCGAC GACGGTCACC GCTATTCAGC ACAGAATGGC CCAACCTGAG TTCTGCTATC AATGGCCTAT

6901 AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA
     TCCGCGTCGC CAGCCCGACT TGCCCCCCAA GCACGTGTGT CGGGTCGAAC CTCGCTTGCT GGATGTGGCT TGACTCTATG GATGTCGCAC TCGATACTCT

7001 AAGCGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCGG AACAGGAGAG CGCACGAGGG AGCTTCCAGG GGGAAACGCC
     TTCGCGGTGC GAAGGGCTTC CCTCTTTCCG CCTGTCCATA GGCCATTCGC CGTCCCAGCC TTGTCCTCTC GCGTGCTCCC TCGAAGGTCC CCCTTTGCGG

7101 TGGTATCTTT ATAGTCCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGTCG ATTTTTGTGA TGCTCGTCAG GGGGGCGGAG CCTATGGAAA AACGCCAGCA
     ACCATAGAAA TATCAGGACA GCCCAAAGCG GTGGAGACTG AACTCGCAGC TAAAAACACT ACGAGCAGTC CCCCCGCCTC GGATACCTTT TTGCGGTCGT
```

FIGURE 5J

```
7201 ACGCGGCCTT TTTACGGTTC CTGGCCTTTT GCTGGCCTTT TGCTCACATG TTCTTTCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCTT
     TGCGCCGGAA AAATGCCAAG GACCGGAAAA CGACCGGAAA ACGAGTGTAC AAGAAAGGAC GCAATAGGGG ACTAAGACAC CTATTGGCAT AATGGCGGAA

7301 TGAGTGAGCT GATACCGCTC GCCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA GAGCGCCCAA TACGCAAACC GCCTCTCCCC
     ACTCACTCGA CTATGGCGAG CGGCGTCGGC TTGCTGGCTC GCGTCGCTCA GTCACTCGCT CCTTCGCCTT CTCGCGGGTT ATGCGTTTGG CGGAGAGGGG

7401 GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGGC AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCATT
     CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGGCTGAC CTTTCGCCCG TCACTCGCGT TGCGTTAATT ACACTCAATC GAGTGAGTAA

7501 AGGCACCCCA GGCTTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAATT TCACACAGGA AACAGCTATG ACATGATTAC
     TCCGTGGGGT CCGAAATGTG AAATACGAAG GCCGAGCATA CAACACACCT TAACACTCGC CTATTGTTAA AGTGTGTCCT TTGTCGATAC TGTACTAATG

7601 GAATTAA
     CTTAATT
```

```
  1 GCGGCCGCAG AGAAAGCAGA GGATGGGGCT TAGCAGCTGG CAGAGCCAGG AGCGGGGAGG TAGCAGAAAG ACCACAAGTA CAAAGAAGTC CTGAAACTTT
    CGCCGGCGTC TCTTTCGTCT CCTACCCCGA ATCGTCGACC GTCTCGGTCC TCGCCCCTCC ATCGTCTTTC TGGTGTTCAT GTTTCTTCAG GACTTTGAAA

101 GGTTTTGCTG CTGCAGCCCA TTGAGAGTGA CGACATGGAG CACAAGACCC TGAAGATCAC CGACTTTGGC CTGGCCCGAG AGTGGCACAA AACCACACAA
    CCAAAACGAC GACGTCGGGT AACTCTCACT GCTGTACCTC GTGTTCTGGG ACTTCTAGTG GCTGAAACCG GACCGGGCTC TCACCGTGTT TTGGTGTGTT

201 ATGAGTGCCG CXGGCACCTA CXCCTGGATG GCTCCTGAGG TTATCAAGGC CTCCACCTTC TCTAAGGGCA GTGACGTCTG GAGTTTTGGG GTGCTGCTGT
    TACTCACGGC G?CCGTGGAT G?GGACCTAC CGAGGACTCC AATAGTTCCG GAGGTGGAAG AGATTCCCGT CACTGCAGAC CTCAAAACCC CACGACGACA

301 GGGAACTGCT GACCGGGGAG XTGCCATACC GTGGCATTGA CTGCCTTGCT GTGGCCTATG GCGTAGCTGT TAACAAGCTC ACACTGCCAT CCATCCACCT
    CCCTTGACGA CTGGCCCCTC ?ACGGTATGG CACCGTAACT GACGGAACGA CACCGGATAC CGCATCGACA ATTGTTCGAG TGTGACGGTA GGTAGGTGGA

401 GGCC
    CCGG
```

FIGURE 6

FIGURE 7A

```
  1 ATGAGACCGT TGGCGGCGGA CGGCGGCCAG CTGCCGCTGC TCGTTGTTTT TTCTGCAATG ATATTTGGGA CTATTACAAA TCAAGATCTG CCTGTGATCA
    TACTCTGGCA ACCGCCGCCT GCCGCCGGTC GACGGCGACG AGCAACAAAA AAGACGTTAC TATAAACCCT GATAATGTTT AGTTCTAGAC GGACACTAGT

101 AGTGTGTTTT AATCAATCAT AAGAACAATG ATTCATCAGT GGGGAAGTCA TCATCATATC CCATGGTATC AGAATCCCCG GAAGACCTCG GGTGTGCGTT
    TCACACAAAA TTAGTTAGTA TTCTTGTTAC TAAGTAGTCA CCCCTTCAGT AGTAGTATAG GGTACCATAG TCTTAGGGGC CTTCTGGAGC CCACACGCAA

201 GAGACCCCAG AGCTCAGGGA CAGTGTACGA AGCTGCCGCT GTGGAAGTGG ATGTATCTGC TTCCATCACA CTGCAAGTGC TGGTCGATGC CCCAGGGAAC
    CTCTGGGGTC TCGAGTCCCT GTCACATGCT TCGACGGCGA CACCTTCACC TACATAGACG AAGGTAGTGT GACGTTCACG ACCAGCTACG GGGTCCCTTG

301 ATTTCCTGTC TCTGGGTCTT TAAGCACAGC TCCCTGAATT GCCAGCCACA TTTTGATTTA CAAAACAGAG GAGTTGTTTC CATGGTCATT TTGAAAATGA
    TAAAGGACAG AGACCCAGAA ATTCGTGTCG AGGGACTTAA CGGTCGGTGT AAAACTAAAT GTTTTGTCTC CTCAACAAAG GTACCAGTAA AACTTTTACT

401 CAGAAACCCA AGCTGGAGAA TACCTACTTT TTATTCAGAG TGAAGCTACC AATTACACAA TATTGTTTAC AGTGAGTATA AGAAATACCC TGCTTTACAC
    GTCTTTGGGT TCGACCTCTT ATGGATGAAA AATAAGTCTC ACTTCGATGG TTAATGTGTT ATAACAAATG TCACTCATAT TCTTTATGGG ACGAAATGTG

501 ATTAAGAAGA CCTTACTTTA GAAAAATGGA AAACCAGGAC GCCCTGGTCT GCATATCTGA GAGCGTTCCA GAGCGGATCC TGGAATGGGT GCTTTGCGAT
    TAATTCTTCT GGAATGAAAT CTTTTTACCT TTTGGTCCTG CGGGACCAGA CGTATAGACT CTCGCAAGGT CTCGCCTAGG ACCTTACCCA CGAAACGCTA

601 TCACAGGGGG AAAGCTGTAA AGAAGAAAGT CCAGCTGTTG TTAAAAAGGA GGAAAAAGTG CTTCATGAAT TATTTGGGAC GGACATAAGG TGCTGTGCCA
    AGTGTCCCCC TTTCGACATT TCTTCTTTCA GGTCGACAAC AATTTTTCCT CCTTTTTCAC GAAGTACTTA ATAAACCCTG CCTGTATTCC ACGACACGGT

701 GAAATGAACT GGGCAGGGAA TGCACCAGGC TGTTCACAAT AGATCTAAAT CAAACTCCTC AGACCACATT GCCACAATTA TTTCTTAAAG TAGGGGAACC
    CTTTACTTGA CCCGTCCCTT ACGTGGTCCG ACAAGTGTTA TCTAGATTTA GTTTGAGGAG TCTGGTGTAA CGGTGTTAAT AAAGAATTTC ATCCCCTTGG
```

FIGURE 7B

```
801  CTTATGGATA AGGTGCAAAG CTGTTCATGT GAACCATGGA TTCGGGCTCA CCTGGGAATT AGAAAACAAA GCACTCGAGG AGGGCAACTA CTTTGAGATG
     GAATACCTAT TCCACGTTTC GACAAGTACA CTTGGTACCT AAGCCCGAGT GGACCCTTAA TCTTTTGTTT CGTGAGCTCC TCCCGTTGAT GAAACTCTAC

901  AGTACCTATT CAACAAACAG AACTATGATA CGGATTCTGT TTGCTTTTGT ATCATCAGTG GCAAGAAACG ACACCGGATA CTACACTTGT TCCTCTTCAA
     TCATGGATAA GTTGTTTGTC TTGATACTAT GCCTAAGACA AACGAAAACA TAGTAGTCAC CGTTCTTTGC TGTGGCCTAT GATGTGAACA AGGAGAAGTT

1001 AGCATCCCAG TCAATCAGCT TTGGTTACCA TCGTAGAAAA GGGATTTATA AATGCTACCA ATTCAAGTGA AGATTATGAA ATTGACCAAT ATGAAGAGTT
     TCGTAGGGTC AGTTAGTCGA AACCAATGGT AGCATCTTTT CCCTAAATAT TTACGATGGT TAAGTTCACT TCTAATACTT TAACTGGTTA TACTTCTCAA

1101 TTGTTTTTCT GTCAGGTTTA AAGCCTACCC ACAAATCAGA TGTACGTGGA CCTTCTCTCG AAAATCATTT CCTTGTGAGC AAAAGGGTCT TGATAACGGA
     AACAAAAAGA CAGTCCAAAT TTCGGATGGG TGTTTAGTCT ACATGCACCT GGAAGAGAGC TTTTAGTAAA GGAACACTCG TTTTCCCAGA ACTATTGCCT

1201 TACAGCATAT CCAAGTTTTG CAATCATAAG CACCAGCCAG GAGAATATAT ATTCCATGCA GAAAATGATG ATGCCCAATT TACCAAAATG TTCACGCTGT
     ATGTCGTATA GGTTCAAAAC GTTAGTATTC GTGGTCGGTC CTCTTATATA TAAGGTACGT CTTTTACTAC TACGGGTTAA ATGGTTTTAC AAGTGCGACA

1301 ATATAAGAAG GAAACCTCAA GTCCTCGGAG AAGCTTCGGC AAGTCAGGCG TCCTGTTTCT CGGATGGATA CCCATTACCA TCTTGGACCT GGAAGAAGTG
     TATATTCTTC CTTTGGAGTT CAGGAGCCTC TTCGAAGCCG TTCAGTCCGC AGGACAAAGA GCCTACCTAT GGGTAATGGT AGAACCTGGA CCTTCTTCAC

1401 TTCAGACAAG TCTCCCAACT GCACAGAAGA GATCACAGAA GGAGTCTGGA ATAGAAAGGC TAACAGAAAA GTGTTTGGAC AGTGGGTGTC GAGCAGTACT
     AAGTCTGTTC AGAGGGTTGA CGTGTCTTCT CTAGTGTCTT CCTCAGACCT TATCTTTCCG ATTGTCTTTT CACAAACCTG TCACCCACAG CTCGTCATGA

1501 CTAAACATGA GTGAAGCCAT AAAAGGGTTC CTGGTCAAGT GCTGTGCATA CAATTCCCTT GGCACATCTT GTGAGACGAT CCTTTTAAAC TCTCCAGGCC
     GATTTGTACT CACTTCGGTA TTTTCCCAAG GACCAGTTCA CGACACGTAT GTTAAGGGAA CCGTGTAGAA CACTCTGCTA GGAAAATTTG AGAGGTCCGG
```

FIGURE 7C

```
1601 CCTTCCCTTT CATCCAAGAC AACATCTCAT TCTATGCAAC AATTGCTGTT TGTCTCCTCT TCATTGTCCT TTTAACCCTG CTAATTTGTC ACAAGTACAA
     GGAAGGGAAA GTAGGTTCTG TTGTAGAGTA AGATACGTTG TTAACGACAA ACAGAGGAGA AGTAACAGGA AAATTGGGAC GATTAAACAG TGTTCATGTT

1701 AAAGCAATTT AGGTATGAAA GCCAGCTACA GATGGTACAG GTGACCGGAT CCTCAGATAA TGAGTACTTC TACGTTGATT TCAGAGAATA TGAATATGAT
     TTTCGTTAAA TCCATACTTT CGGTCGATGT CTACCATGTC CACTGGCCTA GGAGTCTATT ACTCATGAAG ATGCAACTAA AGTCTCTTAT ACTTATACTA

1801 CTCAAATGGG AGTTTCCAAG AGAAAATTTA GAGTTTGGGA AGGTACTAGG ATCAGGTGCT TTTGGAAAAG TGATGAACGC AACAGCTTAT GGAATTAGCA
     GAGTTTACCC TCAAAGGTTC TCTTTTAAAT CTCAAACCCT TCCATGATCC TAGTCCACGA AAACCTTTTC ACTACTTGCG TTGTCGAATA CCTTAATCGT

1901 AAACAGGAGT CTCAATCCAG GTTACCGTCA AAATGCTGAA AGAAAAAGCA GACAGCTCTG AAAGAGAGGC ACTCATGTCA GAACTCAAGA TGATGACCCA
     TTTGTCCTCA GAGTTAGGTC CAATGGCAGT TTTACGACTT TCTTTTTCGT CTGTCGAGAC TTTCTCTCCG TGAGTACAGT CTTGAGTTCT ACTACTGGGT

2001 GCTGGGAAGC CACGAGAATA TTGTGAACCT GCTGGGGGCG TGCACACTGT CAGGACCAAT TTACTTGATT TTTGAATACT GTTGCTATGG TGATCTTCTC
     CGACCCTTCG GTGCTCTTAT AACACTTGGA CGACCCCCGC ACGTGTGACA GTCCTGGTTA AATGAACTAA AAACTTATGA CAACGATACC ACTAGAAGAG

2101 AACTATCTAA GAAGTAAAAG AGAAAAATTT CACAGGACTT GGACAGAGAT TTTCAAGGAA CACAATTTCA GTTTTTACCC CACTTTCCAA TCACATCCAA
     TTGATAGATT CTTCATTTTC TCTTTTTAAA GTGTCCTGAA CCTGTCTCTA AAAGTTCCTT GTGTTAAAGT CAAAAATGGG GTGAAAGGTT AGTGTAGGTT

2201 ATTCCAGCAT GCCTGGTTCA AGAGAAGTTC AGATACACCC GGACTCGGAT CAAATCTCAG GGCTTCATGG GAATTCATTT CACTCTGAAG ATGAAATTGA
     TAAGGTCGTA CGGACCAAGT TCTCTTCAAG TCTATGTGGG CCTGAGCCTA GTTTAGAGTC CCGAAGTACC CTTAAGTAAA GTGAGACTTC TACTTTAACT

2301 ATATGAAAAC CAAAAAAGGC TGGAAGAAGA GGAGGACTTG AATGTGCTTA CATTTGAAGA TCTTCTTTGC TTTGCATATC AAGTTGCCAA AGGAATGGAA
     TATACTTTTG GTTTTTTCCG ACCTTCTTCT CCTCCTGAAC TTACACGAAT GTAAACTTCT AGAAGAAACG AAACGTATAG TTCAACGGTT TCCTTACCTT
```

FIGURE 7D

```
2401 TTTCTGGAAT TTAAGTCGTG TGTTCACAGA GACCTGGCCG CCAGGAACGT GCTTGTCACC CACGGGAAAG TGGTGAAGAT ATGTGACTTT GGATTGGCTC
     AAAGACCTTA AATTCAGCAC ACAAGTGTCT CTGGACCGGC GGTCCTTGCA CGAACAGTGG GTGCCCTTTC ACCACTTCTA TACACTGAAA CCTAACCGAG

2501 GAGATATCAT GAGTGATTCC AACTATGTTG TCAGGGGCAA TGCCCGTCTG CCTGTAAAAT GGATGGCCCC CGAAAGCCTG TTTGAAGGCA TCTACACCAT
     CTCTATAGTA CTCACTAAGG TTGATACAAC AGTCCCCGTT ACGGGCAGAC GGACATTTTA CCTACCGGGG GCTTTCGGAC AAACTTCCGT AGATGTGGTA

2601 TAAGAGTGAT GTCTGGTCAT ATGGAATATT ACTGTGGGAA ATCTTCTCAC TTGGTGTGAA TCCTTACCCT GGCATTCCGG TTGATGCTAA CTTCTACAAA
     ATTCTCACTA CAGACCAGTA TACCTTATAA TGACACCCTT TAGAAGAGTG AACCACACTT AGGAATGGGA CCGTAAGGCC AACTACGATT GAAGATGTTT

2701 CTGATTCAAA ATGGATTTAA AATGGATCAG CCATTTTATG CTACAGAAGA AATATACATT ATAATGCAAT CCTGCTGGGC TTTTGACTCA AGGAAACGGC
     GACTAAGTTT TACCTAAATT TTACCTAGTC GGTAAAATAC GATGTCTTCT TTATATGTAA TATTACGTTA GGACGACCCG AAAACTGAGT TCCTTTGCCG

2801 CATCCTTCCC TAATTTGACT TCGTTTTTAG GATGTCAGCT GGCAGATGCA GAAGAAGCGA TGTATCAGAA TGTGGATGGC CGTGTTTCGG AATGTCCTCA
     GTAGGAAGGG ATTAAACTGA AGCAAAAATC CTACAGTCGA CCGTCTACGT CTTCTTCGCT ACATAGTCTT ACACCTACCG GCACAAAGCC TTACAGGAGT

2901 CACCTACCAA AACAGGCGAC CTTTCAGCAG AGAGATGGAT TTGGGGCTAC TCTCTCCGCA GGCTCAGGTC GAAGATTCGT AGAGGAACAA TTTAGTTTTA
     GTGGATGGTT TTGTCCGCTG GAAAGTCGTC TCTCTACCTA AACCCCGATG AGAGAGGCGT CCGAGTCCAG CTTCTAAGCA TCTCCTTGTT AAATCAAAAT

3001 AGGACTTCAT CCCTCCACCT ATCCCTAACA GGCTGTAGAT TACCAAAACA AGGTTAATTT CATCACTAAA AGAAAATCTA TTATCAACTG CTGCTTCACC
     TCCTGAAGTA GGGAGGTGGA TAGGGATTGT CCGACATCTA ATGGTTTTGT TCCAATTAAA GTAGTGATTT TCTTTTAGAT AATAGTTGAC GACGAAGTGG

3101 AGACTTTTCT CTAGAGAGCG
     TCTGAAAAGA GATCTCTCGC
```

FIGURE 8A

```
  1 TCGGCGTCCA CCCGCCCAGG GAGAGTCAGA CCTGGGGGGG CGAGGGCCCC CCAAACTCAG TTCGGATCCT ACCCGAGTGA GGCGGCGCCA TGGAGCTCCG
    AGCCGCAGGT GGGCGGGTCC CTCTCAGTCT GGACCCCCCC GCTCCCGGGG GGTTTGAGTC AAGCCTAGGA TGGGCTCACT CCGCCGCGGT ACCTCGAGGC
  1                                                                                                  M  E  L  R

101 GGTGCTGCTC TGCTGGGCTT CGTTGGCCGC AGCTTTGGAA GAGACCCTGC TGAACACAAA ATTGGAAACT GCTGATCTGA AGTGGGTGAC ATTCCCTCAG
    CCACGACGAG ACGACCCGAA GCAACCGGCG TCGAAACCTT CTCTGGGACG ACTTGTGTTT TAACCTTTGA CGACTAGACT TCACCCACTG TAAGGGAGTC
  5  V  L  L  C  W  A  S  L  A  A  A  L  E  E  T  L  L  N  T  K  L  E  T  A  D  L  K  W  V  T  F  P  Q

201 GTGGACGGGC AGTGGGAGGA ACTGAGCGGC CTGGATGAGG AACAGCACAG CGTGCGCACC TACGAAGTGT GTGACGTGCA GCGTGCCCCG GGCCAGGCCC
    CACCTGCCCG TCACCCTCCT TGACTCGCCG GACCTACTCC TTGTCGTGTC GCACGCGTGG ATGCTTCACA CACTGCACGT CGCACGGGGC CCGGTCCGGG
 38 V  D  G  Q  W  E  E  L  S  G  L  D  E  E  Q  H  S  V  R  T  Y  E  V  C  D  V  Q  R  A  P  G  Q  A  H

301 ACTGGCTTCG CACAGGTTGG GTCCCACGGC GGGGCGCCGT CCACGTGTAC GCCACGCTGC GCTTCACCAT GCTCGAGTGC CTGTCCCTGC CTCGGGCTGG
    TGACCGAAGC GTGTCCAACC CAGGGTGCCG CCCCGCGGCA GGTGCACATG CGGTGCGACG CGAAGTGGTA CGAGCTCACG GACAGGGACG GAGCCCGACC
 72  W  L  R  T  G  W  V  P  R  R  G  A  V  H  V  Y  A  T  L  R  F  T  M  L  E  C  L  S  L  P  R  A  G

401 GCGCTCCTGC AAGGAGACCT TCACCGTCTT CTACTATGAG AGCGATGCGG ACACGGCCAC GGCCCTCACG CCAGCCTGGA TGGAGAACCC CTACATCAAG
    CGCGAGGACG TTCCTCTGGA AGTGGCAGAA GATGATACTC TCGCTACGCC TGTGCCGGTG CCGGGAGTGC GGTCGGACCT ACCTCTTGGG GATGTAGTTC
105  R  S  C  K  E  T  F  T  V  F  Y  Y  E  S  D  A  D  T  A  T  A  L  T  P  A  W  M  E  N  P  Y  I  K

501 GTGGACACGG TGGCCGCGGA GCATCTCACC CGGAAGCGCC CTGGGGCCGA GGCCACCGGG AAGGTGAATG TCAAGACGCT GCGTCTGGGA CCGCTCAGCA
    CACCTGTGCC ACCGGCGCCT CGTAGAGTGG GCCTTCGCGG GACCCCGGCT CCGGTGGCCC TTCCACTTAC AGTTCTGCGA CGCAGACCCT GGCGAGTCGT
138 V  D  T  V  A  A  E  H  L  T  R  K  R  P  G  A  E  A  T  G  K  V  N  V  K  T  L  R  L  G  P  L  S  K

601 AGGCTGGCTT CTACCTGGCC TTCCAGGACC AGGGTGCCTG CATGGCCCTG CTATCCCTGC ACCTCTTCTA CAAAAAGTGC GCCCAGCTGA CTGTGAACCT
    TCCGACCGAA GATGGACCGG AAGGTCCTGG TCCCACGGAC GTACCGGGAC GATAGGGACG TGGAGAAGAT GTTTTTCACG CGGGTCGACT GACACTTGGA
172  A  G  F  Y  L  A  F  Q  D  Q  G  A  C  M  A  L  L  S  L  H  L  F  Y  K  K  C  A  Q  L  T  V  N  L
```

FIGURE 8B

```
801  TACTGCCGTG AGGATGGCCA GTGGGCCGAA CAGCCGGTCA CGGGCTGCAG CTGTGCTCCG GGGTTCGAGG CAGCTGAGGG GAACACCAAG TGCCGAGCCT
     ATGACGGCAC TCCTACCGGT CACCCGGCTT GTCGGCCAGT GCCCGACGTC GACACGAGGC CCCAAGCTCC GTCGACTCCC CTTGTGGTTC ACGGCTCGGA
210  Y  C  R  E   D  G  Q   W  A  E   Q  P  V  T   G  C  S   C  A  P   G  F  E  A   A  E  G   N  T  K   C  R  A  C

901  GTGCCCAGGG CACCTTCAAG CCCCTGTCAG GAGAAGGGTC CTGCCAGCCA TGCCCAGCCA ATAGCCACTC TAACACCATT GGATCAGCCG TCTGCCAGTG
     CACGGGTCCC GTGGAAGTTC GGGGACAGTC CTCTTCCCAG GACGGTCGGT ACGGGTCGGT TATCGGTGAG ATTGTGGTAA CCTAGTCGGC AGACGGTCAC
272   A  Q  G   T  F  K   P  L  S  G   E  G  S   C  Q  P   C  P  A  N   S  H  S   N  T  I   G  S  A  V   C  Q  C

1001 CCGCGTCGGG TACTTCCGGG CACGCACAGA CCCCCGGGGT GCACCCTGCA CCACCCCTCC TTCGGCTCCG CGGAGCGTGG TTTCCCGCCT GAACGGCTCC
     GGCGCAGCCC ATGAAGGCCC GTGCGTGTCT GGGGGCCCCA CGTGGGACGT GGTGGGGAGG AAGCCGAGGC GCCTCGCACC AAAGGGCGGA CTTGCCGAGG
305   R  V  G   Y  F  R  A   R  T  D   P  R  G   A  P  C  T   T  P  P   S  A  P   R  S  V  V   S  R  L   N  G  S

1101 TCCCTGCACC TGGAATGGAG TGCCCCCCTG GAGTCTGGTG GCCGAGAGGA CCTCACCTAC GCCCTCCGCT GCCGGGAGTG CCGACCCGGA GGCTCCTGTG
     AGGGACGTGG ACCTTACCTC ACGGGGGGAC CTCAGACCAC CGGCTCTCCT GGAGTGGATG CGGGAGGCGA CGGCCCTCAC GGCTGGGCCT CCGAGGACAC
338  S  L  H  L   E  W  S   A  P  L   E  S  G  G   R  E  D   L  T  Y   A  L  R  C   R  E  C   R  P  G   G  S  C  A

1201 CGCCCTGCGG GGGAGACCTG ACTTTTGACC CCGGCCCCCG GGACCTGGTG GAGCCCTGGG TGGTGGTTCG AGGGCTACGT CCTGACTTCA CCTATACCTT
     GCGGGACGCC CCCTCTGGAC TGAAAACTGG GGCCGGGGGC CCTGGACCAC CTCGGGACCC ACCACCAAGC TCCCGATGCA GGACTGAAGT GGATATGGAA
372   P  C  G   G  D  L   T  F  D  P   G  P  R   D  L  V   E  P  W  V   V  V  R   G  L  R   P  D  F  T   Y  T  F

1301 TGAGGTCACT GCATTGAACG GGGTATCCTC CTTAGCCACG GGGCCCGTCC CATTTGAGCC TGTCAATGTC ACCACTGACC GAGAGGTACC TCCTGCAGTG
     ACTCCAGTGA CGTAACTTGC CCCATAGGAG GAATCGGTGC CCCGGGCAGG GTAAACTCGG ACAGTTACAG TGGTGACTGG CTCTCCATGG AGGACGTCAC
405   E  V  T   A  L  N  G   V  S  S   L  A  T   G  P  V  P   F  E  P   V  N  V   T  T  D  R   E  V  P   P  A  V

1401 TCTGACATCC GGGTGACGCG GTCCTCACCC AGCAGCTTGA GCCTGGCCTG GGCTGTTCCC CGGGCACCCA GTGGGGCTGT GCTGGACTAC GAGGTCAAAT
     AGACTGTAGG CCCACTGCGC CAGGAGTGGG TCGTCGAACT CGGACCGGAC CCGACAAGGG GCCCGTGGGT CACCCCGACA CGACCTGATG CTCCAGTTTA
438  S  D  I  R   V  T  R   S  S  P   S  S  L  S   L  A  W   A  V  P   R  A  P  S   G  A  V   L  D  Y   E  V  K  Y
```

FIGURE 8C

```
1501 ACCATGAGAA GGGCGCCGAG GGTCCCAGCA GCGTGCGGTT CCTGAAGACG TCAGAAAACC GGGCAGAGCT GCGGGGGCTG AAGCGGGGAG CCAGCTACCT
     TGGTACTCTT CCCGCGGCTC CCAGGGTCGT CGCACGCCAA GGACTTCTGC AGTCTTTTGG CCCGTCTCGA CGCCCCCGAC TTCGCCCCTC GGTCGATGGA
472   H  E  K   G  A  E   G  P  S  S   V  R  F   L  K  T   S  E  N  R   A  E  L   R  G  L   K  R  G  A   S  Y  L

1601 GGTGCAGGTA CGGGCGCGCT CTGAGGCCGG CTACGGGCCC TTCGGCCAGG AACATCACAG CCAGACCCAA CTGGATGAGA GCGAGGGCTG GCGGGAGCAG
     CCACGTCCAT GCCCGCGCGA GACTCCGGCC GATGCCCGGG AAGCCGGTCC TTGTAGTGTC GGTCTGGGTT GACCTACTCT CGCTCCCGAC CGCCCTCGTC
505   V  Q  V   R  A  R  S   E  A  G   Y  G  P   F  G  Q  E   H  H  S   Q  T  Q   L  D  E  S   E  G  W   R  E  Q

1701 CTGGCCCTGA TTGCGGGCAC GGCAGTCGTG GGTGTGGTCC TGGTCCTGGT GGTCATTGTG GTCGCAGTTC TCTGCCTCAG GAAGCAGAGC AATGGGAGAG
     GACCGGGACT AACGCCCGTG CCGTCAGCAC CCACACCAGG ACCAGGACCA CCAGTAACAC CAGCGTCAAG AGACGGAGTC CTTCGTCTCG TTACCCTCTC
538  L  A  L  I   A  G  T   A  V  V   G  V  V  L   V  L  V   V  I  V   V  A  V  L   C  L  R   K  Q  S   N  G  R  E

1801 AAGCAGAATA TTCGGACAAA CACGGACAGT ATCTCATCGG ACATGGTACT AAGGTCTACA TCGACCCCTT CACTTATGAA GACCCTAATG AGGCTGTGAG
     TTCGTCTTAT AAGCCTGTTT GTGCCTGTCA TAGAGTAGCC TGTACCATGA TTCCAGATGT AGCTGGGGAA GTGAATACTT CTGGGATTAC TCCGACACTC
572   A  E  Y   S  D  K   H  G  Q  Y   L  I  G   H  G  T   K  V  Y  I   D  P  F   T  Y  E   D  P  N  E   A  V  R

1901 GGAATTTGCA AAAGAGATCG ATGTCTCCTA CGTCAAGATT GAAGAGGTGA TTGGTGCAGG TGAGTTTGGC GAGGTGTGCC GGGGGCGGCT CAAGGCCCCA
     CCTTAAACGT TTTCTCTAGC TACAGAGGAT GCAGTTCTAA CTTCTCCACT AACCACGTCC ACTCAAACCG CTCCACACGG CCCCCGCCGA GTTCCGGGGT
605   E  F  A   K  E  I  D   V  S  Y   V  K  I   E  E  V  I   G  A  G   E  F  G   E  V  C  R   G  R  L   K  A  P

2001 GGGAAGAAGG AGAGCTGTGT GGCAATCAAG ACCCTGAAGG GTGGCTACAC GGAGCGGCAG CGGCGTGAGT TTCTGAGCGA GGCCTCCATC ATGGGCCAGT
     CCCTTCTTCC TCTCGACACA CCGTTAGTTC TGGGACTTCC CACCGATGTG CCTCGCCGTC GCCGCACTCA AAGACTCGCT CCGGAGGTAG TACCCGGTCA
638  G  K  K  E   S  C  V   A  I  K   T  L  K  G   G  Y  T   E  R  Q   R  R  E  F   L  S  E   A  S  I   M  G  Q  F

2101 TCGAGCACCC CAATATCATC CGCCTGGAGG GCGTGGTCAC CAACAGCATG CCCGTCATGA TTCTCACAGA GTTCATGGAG AACGGCGCCC TGGACTCCTT
     AGCTCGTGGG GTTATAGTAG GCGGACCTCC CGCACCAGTG GTTGTCGTAC GGGCAGTACT AAGAGTGTCT CAAGTACCTC TTGCCGCGGG ACCTGAGGAA
672   E  H  P   N  I  I   R  L  E  G   V  V  T   N  S  M   P  V  M  I   L  T  E   F  M  E   N  G  A  L   D  S  F
```

FIGURE 8D

```
2201 CCTGCGGCTA AACGACGGAC AGTTCACAGT CATCCAGCTC GTGGGCATGC TGCGGGGCAT CGCCTCGGGC ATGCGGTACC TTGCCGAGAT GAGCTACGTC
     GGACGCCGAT TTGCTGCCTG TCAAGTGTCA GTAGGTCGAG CACCCGTACG ACGCCCCGTA GCGGAGCCCG TACGCCATGG AACGGCTCTA CTCGATGCAG
705   L  R  L   N  D  G  Q   F  T  V   I  Q  L   V  G  M  L   R  G  I   A  S  G   M  R  Y  L   A  E  M   S  Y  V

2301 CACCGAGACC TGGCTGCTCG CAACATCCTA GTCAACAGCA ACCTCGTCTG CAAAGTGTCT GACTTTGGCC TTTCCCGATT CCTGGAGGAG AACTCTTCCG
     GTGGCTCTGG ACCGACGAGC GTTGTAGGAT CAGTTGTCGT TGGAGCAGAC GTTTCACAGA CTGAAACCGG AAAGGGCTAA GGACCTCCTC TTGAGAAGGC
738  H  R  D  L   A  A  R   N  I  L   V  N  S  N   L  V  C   K  V  S   D  F  G  L   S  R  F   L  E  E   N  S  S  D

2401 ATCCCACCTA CACGAGCTCC CTGGGAGGAA AGATTCCCAT CCGATGGACT GCCCCGGAGG CCATTGCCTT CCGGAAGTTC ACTTCCGCCA GTGATGCCTG
     TAGGGTGGAT GTGCTCGAGG GACCCTCCTT TCTAAGGGTA GGCTACCTGA CGGGGCCTCC GGTAACGGAA GGCCTTCAAG TGAAGGCGGT CACTACGGAC
772   P  T  Y   T  S  S   L  G  G  K   I  P  I   R  W  T   A  P  E  A   I  A  F   R  K  F   T  S  A  S   D  A  W

2501 GAGTTACGGG ATTGTGATGT GGGAGGTGAT GTCATTTGGG GAGAGGCCGT ACTGGGACAT GAGCAATCAG GACGTGATCA ATGCCATTGA ACAGGACTAC
     CTCAATGCCC TAACACTACA CCCTCCACTA CAGTAAACCC CTCTCCGGCA TGACCCTGTA CTCGTTAGTC CTGCACTAGT TACGGTAACT TGTCCTGATG
805   S  Y  G   I  V  M  W   E  V  M   S  F  G   E  R  P  Y   W  D  M   S  N  Q   D  V  I  N   A  I  E   Q  D  Y

2601 CGGCTGCCCC CGCCCCCAGA CTGTCCCACC TCCCTCCACC AGCTCATGCT GGACTGTTGG CAGAAAGACC GGAATGCCCG GCCCCGCTTC CCCCAGGTGG
     GCCGACGGGG GCGGGGGTCT GACAGGGTGG AGGGAGGTGG TCGAGTACGA CCTGACAACC GTCTTTCTGG CCTTACGGGC CGGGGCGAAG GGGGTCCACC
838  R  L  P  P   P  P  D   C  P  T   S  L  H  Q   L  M  L   D  C  W   Q  K  D  R   N  A  R   P  R  F   P  Q  V  V

2701 TCAGCGCCCT GGACAAGATG ATCCGGAACC CCGCCAGCCT CAAAATCGTG GCCCGGGAGA ATGGCGGGGC CTCACACCCT CTCCTGGACC AGCGGCAGCC
     AGTCGCGGGA CCTGTTCTAC TAGGCCTTGG GGCGGTCGGA GTTTTAGCAC CGGGCCCTCT TACCGCCCCG GAGTGTGGGA GAGGACCTGG TCGCCGTCGG
872   S  A  L   D  K  M   I  R  N  P   A  S  L   K  I  V   A  R  E  N   G  G  A   S  H  P   L  L  D  Q   R  Q  P

2801 TCACTACTCA GCTTTTGGCT CTGTGGGCGA GTGGCTTCGG GCCATCAAAA TGGGAAGATA CGAAGAAAGT TTCGCAGCCG CTGGCTTTGG CTCCTTCGAG
     AGTGATGAGT CGAAAACCGA GACACCCGCT CACCGAAGCC CGGTAGTTTT ACCCTTCTAT GCTTCTTTCA AAGCGTCGGC GACCGAAACC GAGGAAGCTC
905   H  Y  S   A  F  G  S   V  G  E   W  L  R   A  I  K  M   G  R  Y   E  E  S   F  A  A  A   G  F  G   S  F  E
```

FIGURE 8E

```
2901 CTGGTCAGCC AGATCTCTGC TGAGGACCTG CTCCGAATCG GAGTCACTCT GGCGGGACAC CAGAAGAAAA TCTTGGCCAG TGTCCAGCAC ATGAAGTCCC
     GACCAGTCGG TCTAGAGACG ACTCCTGGAC GAGGCTTAGC CTCAGTGAGA CCGCCCTGTG GTCTTCTTTT AGAACCGGTC ACAGGTCGTG TACTTCAGGG
 939 L  V  S  Q  I  S  A  E  D  L  L  R  I  G  V  T  L  A  G  H  Q  K  K  I  L  A  S  V  Q  H  M  K  S  Q

3001 AGGCCAAGCC GGGAACCCCG GGTGGGACAG GAGGACCGGC CCCGCAGTAC TGACCTGCAG GAACTCCCCA CCCCAGGGAC ACCCCCTCCC CATTTTCCGG
     TCCGGTTCGG CCCTTGGGGC CCACCCTGTC CTCCTGGCCG GGGCGTCATG ACTGGACGTC CTTGAGGGGT GGGGTCCCTG TGGGGGAGGG GTAAAAGGCC
 972 A  K  P  G  T  P  G  G  T  G  G  P  A  P  Q  Y  Q  P  A  G  T  P  H  P  R  D  T  A  S  P  F  S  G

3101 GGCAGAGTGG GGACTCACAG AGGCCCCCAG CCCTGTGCCC CGCTGGATTG CACTTTGAGC CCGTGGGGTG AGGAGTTGGC AATTTGGAGA GACAGGATTT
     CCGTCTCACC CCTGAGTGTC TCCGGGGGTC GGGACACGGG GCGACCTAAC GTGAAACTCG GGCACCCCAC TCCTCAACCG TTAAACCTCT CTGTCCTAAA
1005 A  E  W  G  L  T  E  A  P  S  P  V  P  R  W  I  A  L  Q  A  R  G  V  R  S  W  Q  F  G  E  T  G  F

3201 GGGGGTTCTG CCATAATAGG AGGGGAAAAT CACCCCCCAG CCACCTCGGG GAACTCCAGA CCAAGGGTGA GGGCGCCTTT CCCTCAGGAC TGGGTGTGAC
     CCCCCAAGAC GGTATTATCC TCCCCTTTTA GTGGGGGGTC GGTGGAGCCC CTTGAGGTCT GGTTCCCACT CCCGCGGAAA GGGAGTCCTG ACCCACACTG
1038 G  G  S  A  I  I  G  G  E  N  H  P  P  A  T  S  G  N  S  R  P  R  V  R  A  P  F  P  Q  D  W  V  Q  P

3301 CAGAGGAAAA GGAAGTGCCC AACATCTCCC AGCCTCCCCA GGTGCCCCCC TCACCTTGAT GGGTGCGTTC CCGCAGACCA AAGAGAGTGT GACTCCCTTG
     GTCTCCTTTT CCTTCACGGG TTGTAGAGGG TCGGAGGGGT CCACGGGGGG AGTGGAACTA CCCACGCAAG GGCGTCTGGT TTCTCTCACA CTGAGGGAAC
1072 E  E  K  E  V  P  N  I  S  Q  P  P  Q  V  P  P  S  P  Q  W  V  R  S  R  R  P  K  R  V  Q  L  P  C

3401 CCAGCTCCAG AGTGGGGGGG CTGTCCCAGG GGGCAAGAAG GGGTGTCAGG GCCCAGTGAC AAAATCATTG GGGTTTGTAG TCCCAACTTG CTGCTGTCAC
     GGTCGAGGTC TCACCCCCCC GACAGGGTCC CCCGTTCTTC CCCACAGTCC CGGGTCACTG TTTTAGTAAC CCCAAACATC AGGGTTGAAC GACGACAGTG
1105 Q  L  Q  S  G  G  A  V  P  G  G  K  K  G  C  Q  G  P  V  T  K  S  L  G  F  V  V  P  T  C  C  H

3501 CACCAAACTC AATCATTTTT TTCCCTTGTA AATGCCCCTC CCCCAGCTGC TGCCTTCATA TTGAAGGTTT TTGAGTTTTG TTTTTGGTCT TAATTTTTCT
     GTGGTTTGAG TTAGTAAAAA AAGGGAACAT TTACGGGGAG GGGGTCGACG ACGGAAGTAT AACTTCCAAA AACTCAAAAC AAAAACCAGA ATTAAAAAGA
1138 H  Q  T  Q  S  F  F  S  L  V  N  A  P  P  P  A  A  A  F  I  L  K  V  F  E  F  C  P  W  S  Q  F  F  S
```

FIGURE 8F

```
3601 CCCCGTTCCC TTTTTGTTTC TTCGTTTTGT TTTTCTACCG TCCTTGTCAT AACTTTGTGT TGGAGGGAAC CTGTTTCACT ATGGCCTCCT TTGCCCAAGT
     GGGGCAAGGG AAAAACAAAG AAGCAAAACA AAAAGATGGC AGGAACAGTA TTGAAACACA ACCTCCCTTG GACAAAGTGA TACCGGAGGA AACGGGTTCA
1172   P  F  P   F  C  F   F  V  L  F   F  Y  R   P  C  H   N  F  V  L   E  G  T   C  P  T   M  A  S  F   A  Q  V

3701 TGAAACAGGG GCCCATCATC ATGTCTGTTT CCAGAACAGT GCCTTGGTCA TCCCACATCC CCGGACCCCG CCTGGGACCC CCAAGCTGTG TCCTATGAAG
     ACTTTGTCCC CGGGTAGTAG TACAGACAAA GGTCTTGTCA CGGAACCAGT AGGGTGTAGG GGCCTGGGGC GGACCCTGGG GGTTCGACAC AGGATACTTC
1205   E  T  G   A  H  H  H   V  C  F   Q  N  S   A  L  V  I   P  H  P   R  T  P   P  G  T  P   K  L  C   P  H  K

3801 GGGTGTGGGG TGAGGTAGTG AAAAGGGCGG TAGTTGGTGG TGGAACCCAG AAACGGACGC CGGTGCTTGG AGGGGTTCTT AAATTATATT TAAAAAAGTA
     CCCACACCCC ACTCCATCAC TTTTCCCGCC ATCAACCACC ACCTTGGGTC TTTGCCTGCG GCCACGAACC TCCCCAAGAA TTTAATATAA ATTTTTTCAT
1218 G  C  G  V   R  O  O   K  G  R   O  L  V  V   E  P  R   N  G  R   R  C  L  E   G  F  L   N  Y  I   O  K  S  N

3901 ACTTTTTGTA TAAATAAAAG AAAATGGGAC GTGTCCCAGC TCCAGGGGTA AAAAAAAAAA AAAAAAAAA
     TGAAAAACAT ATTTATTTTC TTTTACCCTG CACAGGGTCG AGGTCCCCAT TTTTTTTTTT TTTTTTTTT
1272   F  L  Y   K  O  K   K  M  G  R   V  P  A   P  G  V   K  K  K  K   K  R
```

FIGURE 9	ARNILVNSNLVCKVSDFGLSRFLEDDTSDPTYTSALGGKIPMRWTAPEAIQYRKFASAS
FIGURE 10	NVLVKSPNHVKITDFGLARLLEGDEKEYNADGGKMPIKWMALECIHYRKFTHQS
FIGURE 11	NCMLAGDMTVCVADFGLSWKIYSGATIVRGCASKLPVKWLALGSLADNLYTVHS
FIGURE 12	NCLVGKNYTIKIADFGMSRNLYSGDYY
FIGURE 13	TRNILVENENRVKIGDFGLTKVLPQDKEYYKVKEPGESPIFWYAPESLTESLFSVASD
FIGURE 14	ARNILVNSNLVCKVSDFGMSRVLEDDPEAAYTTRGGKIPIRWTAPEAIAYRKFTSASD

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 93/00586

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/12; C12N15/54; C12N9/12; //C12Q1/68, C12N15/11

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12N ; C12Q ; C07K

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
A	NEURON vol. 6, no. 5, May 1991, pages 691 - 704 LAI, C. & LEMKE, G. 'An extended family of protein-tyrosine kinase genes differentially expressed in the verterbrate nervous system' see the whole document ---	1-7

-/--

° Special categories of cited documents : [10]

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

2

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
07 JUNE 1993	02 -07- 1993
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	ANDRES S.M.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 88, no. 23, 1 December 1991, WASHINGTON US pages 10411 - 10415 HOLTRICH, U. ET AL. 'Two additional protein-tyrosine kinases expressed in human lung: fourth member of the fibroblast growth factor receptor family and an intracellular protein-tyrosine kinase' ---	1-7, 13-15
A	GENE vol. 110, no. 2, 15 January 1992, AMSTERDAM NL pages 205 - 211 BRÄUNINGER, A. ET AL. 'Isolation and characterization of a human gene that encodes a new subclass of protein tyrosine kinases' see the whole document ---	1-7, 13-15
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 87, November 1990, WASHINGTON US pages 8913 - 8917 PARTANEN, J. ET AL. 'Putative tyrosine kinases expressed in K-562 human leukemia cells' see especially clone JTK10 ---	1-7
A	MOLECULAR AND CELLULAR BIOLOGY vol. 9, no. 4, April 1989, WASHINGTON US pages 1587 - 1593 HAO, Q.-L. ET AL. 'Isolation and sequence analysis of a novel human tyrosine kinase gene' see the whole document ---	1-7, 16-18
P,A	WO,A,9 214 748 (AMERICAN CYANAMID COMPANY) 3 September 1992 see the whole document -----	1-12